CODEN: JASMAN

ISSN: 0001-4966

The Journal of the Acoustical Society of America

Vol. 112, No. 1 July 2002

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 112, NO. 1, JULY 2002

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 112, NO. 1, JULY 2002

CONTENTS—*Continued from preceding page*

SPEECH PERCEPTION [71]

BIOACOUSTICS [80]

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.*

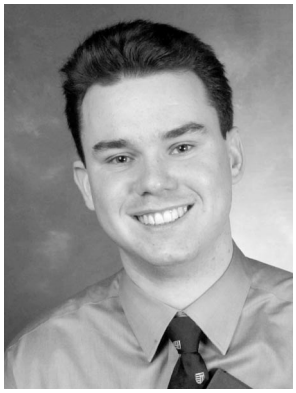
New Fellow of the Acoustical Society of America

David C. Swanson—For contributions to adaptive signal processing in acoustics.

Regional Chapter News

Ashley Woodall with her award-winning acoustics project.

F. V. Hunt Postdoctoral Research Fellowship awarded to Constantine Coussios



Constantine Coussios

The 2002–03 F. V. Hunt Postdoctoral Research Fellowship in Acoustics was awarded to Constantine Coussios. Under the fellowship, Dr. Coussios will undertake a research program at Boston University. The subject of his research is on the modeling of the onset and evolution of a cavitation bubble field induced by high-intensity focused ultrasound in tissue mimicking media. Dr. Coussios holds B.A., M.Eng., M.A. (honorary), and Ph.D. degrees from the University of Cambridge. His Ph.D. thesis was on ''Monitoring of hemolysis by acoustic streaming.'' The Hunt Fellowship is granted each year to an ASA member who has recently received his or her doctorate or will be receiving the degree in the year in which the fellowship is to be granted. The recipient of the fellowship is that individual who, through personal qualifications and a proposed research topic, is judged to exhibit the highest potential for benefiting any aspect of the science of sound and promoting its usefulness to society. Further information about the fellowship is available from the Acoustical Society of America, Suite1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. Phone: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; Web: asa.aip.org/fellowships.html

North Texas: ASA North Texas Chapter presents awards at 45th Dallas Morning News–Toyota Regional Science and Engineering Fair.

More than nine hundred students from 87 North Texas schools competed in the 45th Dallas Morning News–Toyota Regional Science and Engineering Fair. This year, Junior Division contestants won both Outstanding Acoustics Project Awards.

Ms. Jenna Schmear, under the guidance of Ms. Hope May, Blalack Middle School, Carrollton ISD, entered one Outstanding Acoustics Project. Ms. Schmear investigated the sound-attenuating properties of three inexpensive concrete additives. She constructed a sound booth, poured concrete slabs, and, using recorded traffic noise, demonstrated that slabs with additives attenuated traffic noise across a range of intensities more effectively than the all-concrete barrier. Two of the additive slabs use recycled materials; all were less expensive and lighter in weight than barriers currently deployed by her municipal government. Ms. Schmear subsequently won Third Place in the Junior Engineering Division.

Ms. Ashley Woodall, under the guidance of Ms. Janice Borland, Ms. Ginny Lester, and Ms. Margaret Wis, Austin Academy for Excellence, Garland ISD, also entered an Outstanding Acoustics Project. Ms. Woodall investigated the resonant properties of five woods as sustainable replacements for rosewood in marimba bars. She shaped bars from the six woods and then compared their acoustic signatures when struck to find a match for rosewood. Ms. Woodall subsequently won First Place in the Junior Physics Division and Grand Prize in the Junior Division. At the Texas State Science and Engineering Fair, she also won First Place in the Junior Physics Division and Grand Prize in the Junior Division.

Peter Assmann, Ben Seep, Laurie Bornstein, and Mike Daly represented the Chapter.

 0001-4966/2002/112(1)/1/2/$19.00

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.

2002

12–14 August	Recreational Noise Symposium, Key Largo, FL [WWW: http://www.internoise2002.org].
19–21 August	INTER-NOISE 2002, Dearborn, MI [INTER-NOISE 02 Secretariat, The Ohio State University, Department of Mechanical Engineering, 206 West 18th Ave., Columbus, OH 43210-1107, or E-mail: hp@internoise2002.org].
21–25 August	International Hearing Aid Research Conference (IHCON), Lake Tahoe, CA [House Ear Inst., 2100 W. 3rd St., Los Angeles, CA 90057; Fax: 213-413-0950; WWW: http://www.hei.org].
16–20 Sept.	7th International Conference on Spoken Lanugage Processing (ICSLP 2002) Interspeech 2002 [http://www.icslp2002.org/home.html].
2–6 December	First Pan-American/Iberian Meeting on Acoustics (Joint Meeting: 144th Meeting of the Acoustical Society of America, 3rd Iberoamerican Congress of Acoustics, and 9th Mexican Congress on Acoustics), Cancun, Mexico [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org/cancun/cancun.html].

2003

28 April–2 May	145th Meeting of the Acoustical Society of America, Nashville, TN [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
5–8 May	SAE Noise & Vibration Conference & Exhibition, Traverse City, MI [SAE International, 755 W. Big Beaver Rd., Suite 1600, Troy, MI 48084; Fax: 724-776-1830; WWW: http://www.sae.org].
10–14 Nov.	146th Meeting of the Acoustical Society of America, Austin, TX [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

2004

24–28 May	75th Anniversary Meeting (147th Meeting) of the Acoustical Society of America, New York, NY [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
22–26 Nov.	148th Meeting of the Acoustical Society of America, San Diego, CA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

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

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

Some indexes are out of print as noted below.

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

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

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).

OBITUARIES

Evan Mark Relkin • 1951–2002

Evan M. Relkin, a faculty member at the Institute for Sensory Research at Syracuse University, died on 23 March 2002 following a skating accident. Dr. Relkin was a long time member of the Acoustical Society, publishing in the area of physiological acoustics and serving as a member of the Technical Committee on Psychological and Physiological Acoustics.

Dr. Relkin was born 1 October 1951 in New York City, and obtained his undergraduate degree in electrical engineering from the University of Pennsylvania in 1973. He continued studies at the University of Pennsylvania in Bioengineering, and received the M.S.E. in 1975, and continued on with doctoral work. Under the direction of his advisor, James Saunders, Relkin traced the development of the acoustic and mechanical properties of the middle ear in hamsters. This was important work at the time, as the research community in physiological acoustics was attempting to determine the sites and causes of changes in hearing characteristics that occurred in the early life of mammals, and Relkin's work showed that an important source of this change was the middle ear.

Following the receipt of his doctorate in 1979, Relkin spent the next four years as a Research Associate at the Auditory Research Laboratory at Northwestern University, where he worked with Peter Dallos, Jon Siegel, David Harris, and others. In 1983 he accepted a faculty position at Syracuse University, with appointments in the Department of Bioengineering and in the Institute for Sensory Research. Dr. Relkin served as Director of the Department of Bioengineering there from 1984 until 1986. It was during those years in the 1980s that he met his wife, Beth Prieve, who shared with him an interest in auditory function and in the outdoors.

At Syracuse, Dr. Relkin's research was primarily focused on the electrophysiology of the auditory nerve. In collaboration with Denis Pelli at the Institute for Sensory Research, he produced a method for applying signal detection methods to data obtained from auditory nerve spike recordings. This allowed him, in the next few years following the development of the method, to determine (1) that the auditory nerve played only a very small role in behaviorally observed forward masking, and (2) that low-spontaneous rate neurons had a much slower recovery time than high-spontaneous rate auditory neurons. These and other findings were put to further use in collaboration with other researchers at Syracuse, including Jozef Zwislocki, Robert Smith, John Doucet, and Fan-Gang Zeng.

He is survived by his wife Beth, and their two children, Lucas and Elena, who reside in Jamesville, NY, as well as his mother, Lillian Brookman, and brothers Paul and Robert.

CHRISTOPHER W. TURNER

 0001-4966/2002/112(1)/12/1/$19.00

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, 348 Benedum Hall, Pittsburgh, Pennsylvania 15261*
DAVID PREVES, *4 Deerfield Drive, Princeton Junction, New Jersey 08550*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
CARL J. ROSENBERG, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
KEVIN P. SHEPHERD, *Mail Stop 463, NASA Langley Research Center, Hampton, Virginia 23681*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*

6,327,896

43.10.Pr PHOTO-ACOUSTIC LEAK DETECTION SYSTEM

William A. Veronesi *et al.*, assignors to United Technologies Corporation
11 December 2001 (Class 73/40.5A); filed 18 September 1998

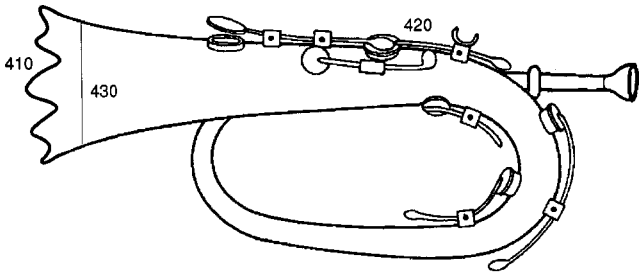
Detection of a gas leak is accomplished by directing a laser beam with a wavelength that the gas absorbs at a point on the component to be monitored. The absorbed energy heats the gas, causing it to produce acoustic noise, which is measured conventionally. In order to distinguish this noise from noise that may be radiated from the test component itself due to its absorption of the laser beam's energy, the test point is scanned by a second laser beam at a wavelength that the gas does not absorb. The corresponding baseline noise signal is subtracted electronically from that obtained with the first laser beam. The same point on the test component is exposed sequentially to the two laser beams, which are directed at that point and switched by an acousto-optic cell.—EEU

6,334,505

43.20.El OPTIMUM EDGES FOR SPEAKERS AND MUSICAL INSTRUMENTS

Ming-Chiang Li, Mitchellville, Maryland
1 January 2002 (Class 181/192); filed 15 November 1990

The use of serrated edges to reduce diffraction is well known in the antenna literature—Here it is applied to a variety of acoustical sources.



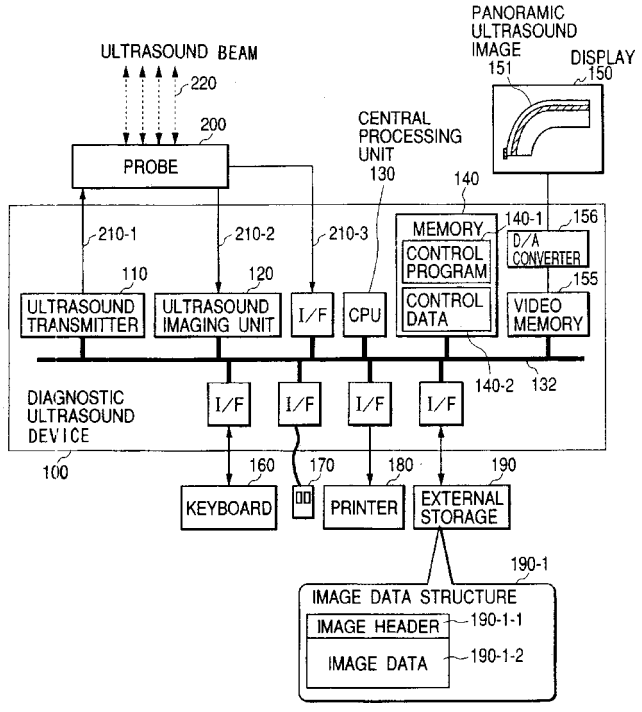
''The musical tone emitted by loudpeakers, brass and wind instruments becomes more smooth, mellow, rich, clean and elegant.''—MK

6,328,693

43.35.Sx METHOD AND DEVICE OF SYNTHESIZING PANORAMA IMAGE FROM ULTRASOUND IMAGES

Takafumi Miyatake *et al.*, assignors to Hitachi, Limited
11 December 2001 (Class 600/437); filed in Japan 28 January 1999

This is a diagnosing ultrasound device that is said to be capable of forming a panoramic slice image in real time. The device converts ultrasound data obtained from a probe into an image, detects shifts between



images formed by the probe, detects rotational effects between sliced images generated by the probe, and displays panoramic images obtained by connecting the sequence of sliced images captured during diagnosis.—DRR

 0001-4966/2002/112(1)/13/12/$19.00

6,334,846

43.35.Wa ULTRASOUND THERAPEUTIC APPARATUS

Yoshiharu Ishibashi *et al.*, assignors to Kabushiki Kaisha Toshiba
1 January 2002 (Class 600/439); filed in Japan 31 March 1995

This apparatus consists of a therapeutic ultrasonic wave generating source and an *in vivo* imaging probe which produces a tissue tomographic image in the vicinity of the ultrasonic wave focus. The imaging probe receives echoes of the pulses emitted from the wave-generating source. The source driving condition is adjusted on the basis of a received echo signal containing information about the ultrasonic wave intensity within a living body, which may lead to increased safety and reliability of the therapy.—DRR

6,328,694

43.35.Yb ULTRASOUND APPARATUS AND METHOD FOR TISSUE RESONANCE ANALYSIS

David Michaeli, assignor to Inta-Medics, Limited
11 December 2001 (Class 600/438); filed 26 May 2000

An ultrasonic probe is placed on a patient's head and is used to generate an ultrasonic pulse, which propagates through the skull and brain and is reflected from the skull and soft tissue lying in a path perpendicular to the probe. The reflected signals are received by the probe and then processed in a specific manner to generate an echo encephalogram (Echo EG) signal that is plotted as a function of amplitude versus distance. A portion of the Echo EG signal is then selected and integrated to generate an echo pulsograph (EPG) signal. An electrocardiograph (ECG) signal for the patient is also generated. Using the ECG signal as a reference, the EPG signal is applied to provide information on the physiological state of the tissue at a depth from the ultrasonic probe corresponding to the selected portion of the Echo EG signal.—DRR

6,328,695

43.35.Yb METHOD AND AN APPARATUS FOR INVESTIGATING MATERIAL PROPERTIES OF BONE USING ULTRASOUND

Klaus Vammen, Birkerdød, Denmark *et al.*
11 December 2001 (Class 600/442); filed in Denmark 22 December 1997

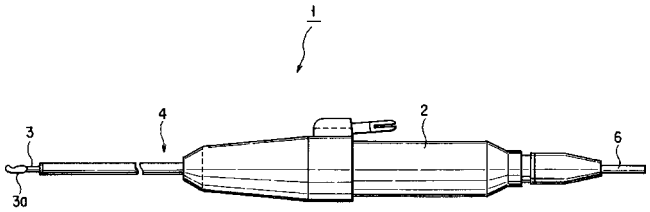
This device is intended to investigate the mechanical properties of bone inside a live animal or human being. It launches an ultrasound pulse into the body of the subject and establishes a trace related to the magnitude of a reflected wave versus the time lapsed since the launching of the pulse. An interval is identified on the trace in which the magnitude of the reflected ultrasound wave exhibits a steady decline versus time. The attenuation of the sound wave within this interval is computed from the trace readings. Osteoporosis is one condition that may be detected by the device.—DRR

6,328,703

43.35.Yb ULTRASONIC TREATING APPARATUS

Eiji Murakami, assignor to Olympus Optical Company, Limited
11 December 2001 (Class 601/4); filed in Japan 2 July 1999

This is a probelike device that consists of an ultrasonic vibrator, a vibrating transmitting member connected to the proximal end of the vibrator, and a treatment section that relays the generated signal into the body.—DRR



6,336,899

43.35.Yb ULTRASONIC DIAGNOSIS APPARATUS

Nobuo Yamazaki, assignor to Kabushiki Kaisha Toshiba
8 January 2002 (Class 600/443); filed in Japan 14 October 1998

This device, labeled with a most generic appellation, is designed to guide a paracentetic needle at the time of using ultrasonic paracentisis, which may be done for biopsy purposes or treatment of tumors or lesions. The apparatus includes a two-dimensional array of ultrasonic transducers. The beams scan a target object three-dimensionally. The transducers receive the ultrasonic echo signals and generate real-time, three-dimensional data on the structural information and blood flow in the object. On the basis of the image data obtained, the device displays navigational information needed to guide a puncture needle toward the target.—DRR

6,326,213

43.35.Zc ACOUSTIC STANDING-WAVE ENHANCEMENT OF A FIBER-OPTIC SALMONELLA BIOSENSOR

Stephen V. Letcher *et al.*, assignors to The Boards of Govenors for Higher Education, State of Rhode Island and Providence Plantations
4 December 2001 (Class 436/518); filed 9 February 1998

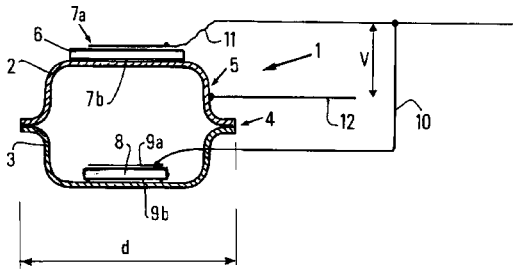
A standing ultrasonic wave is used to concentrate particles to be examined at specific locations along an optical fiber in a test cell, so that the particles can be assayed more precisely.—EEU

6,330,206

43.38.Fx HYDROPHONE FOR ACOUSTIC OR SEISMIC WAVE RECEPTION

Jean Laurent and Georges Constantinou, assignors to Institut Francais du Petrole; Vinci Technologies
11 December 2001 (Class 367/173); filed in France 26 April 1999

A hydrophone comprises two cup-shaped metallic parts **2** and **3** which are joined at their rims **4** to form a sealed housing that supports two different sized piezoelectric discs **6** and **8**, one of which is mounted on the external side of the housing while the other is internal. Because of this unsymmetri-



cal arrangement, the capacitance and sensitivity of the hydrophone are both substantially independent of the operating depth. The unit is potted in acoustically transparent material for subsequent insertion into a tubular shell.—WT

6,320,239

43.38.Gy SURFACE MICROMACHINED ULTRASONIC TRANSDUCER

Peter-Christian Eccardt *et al.*, assignors to Siemens Aktiengesellschaft
20 November 2001 (Class 257/415); filed in Germany 30 October 1996

This patent describes the construction of an ultrasonic transducer by processes like those used in the production of microelectronics, so as to integrate the transducer and its electronics on a chip. A thin layer, for example, of polysilicon, to be used as the diaphragm, is formed atop an auxiliary layer, which is etched away to leave a void between it and the substrate. In order to provide electrostatic excitation of the diaphragm it is made electrically conductive by doping it or by applying a conductive layer atop it, and an electrically conductive region is formed in the substrate by doping. Use of this manufacturing process enables the construction of a micromachined array that can be driven as a phased array.—EEU

6,317,501

43.38.Hz MICROPHONE ARRAY APPARATUS

Naoshi Matsuo, assignor to Fujitsu Limited
13 November 2001 (Class 381/92); filed in Japan 26 June 1997

A number of previous patents address the use of a microphone array to steer the camera of a videoconference system. The digital processing scheme described here is said to be ''...capable of stably and precisely suppressing noise, emphasizing a target sound and identifying the position of a sound source.'' Filter coefficients can be continuously updated even when a talker's voice and unwanted noise are concurrently picked up by the microphones.—GLA

6,320,967

43.38.Ja PASSENGER VEHICLES INCORPORATING LOUDSPEAKERS COMPRISING PANEL-FORM ACOUSTIC RADIATING ELEMENTS

Henry Azima *et al.*, assignors to New Tranducers Limited
20 November 2001 (Class 381/86); filed in the United Kingdom 2 September 1995

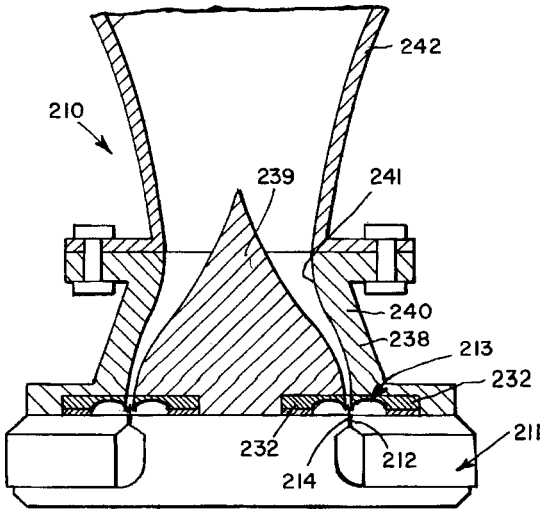
This patent proposes that ''distributed mode acoustic radiators'' be incorporated into automobile interiors. These highly publicized flat loudspeakers consist of a core such as honeycomb or metal foil, and face sheets such as fiberglass or Kevlar reinforced plastic, yielding a composite structure having a high bending stiffness to surface density ratio. Placement locations for these devices include doors, seat backs, and the rear deck.—KPS

6,320,970

43.38.Ja HIGH FREQUENCY COMPRESSION DRIVERS

Eugene J. Czerwinski and Alexander G. Voishvillo, both of Simi Valley, California
20 November 2001 (Class 381/343); filed 25 September 1998

This ''ring radiator'' driver differs from earlier designs in two important respects: First, doubly-curved diaphragm **213** has no distinct compliance portions; it is its own suspension. Second, voice coil support **214** is



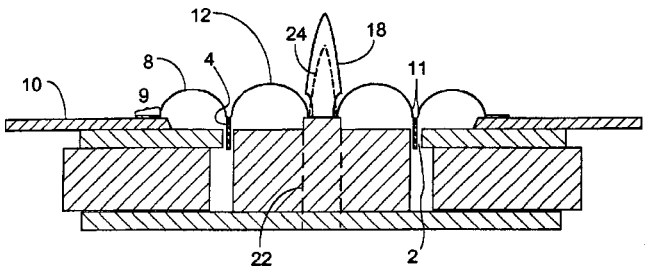
formed as an integral part of the diaphragm structure rather than being glued on. The patent document includes interesting information about general design principles and prior art.—GLA

6,320,972

43.38.Ja LOUDSPEAKER

Lars Goller, assignor to Vifa-Speak A/S
20 November 2001 (Class 381/430); filed in Denmark 17 February 1999

This doubly-curved diaphragm ring radiator might be described as a hornless version of United States Patent 6,320,970, reviewed above. Its novel feature is a projecting central waveguide **18**, which minimizes inter-



ference between forward radiation and centrally reflected sound waves.—GLA

6,324,052

43.38.Ja PERSONAL COMPUTING DEVICES COMPRISING A RESONANT PANEL LOUDSPEAKER

Henry Azima *et al.*, assignors to New Transducers Limited
27 November 2001 (Class 361/683); filed 2 September 1996

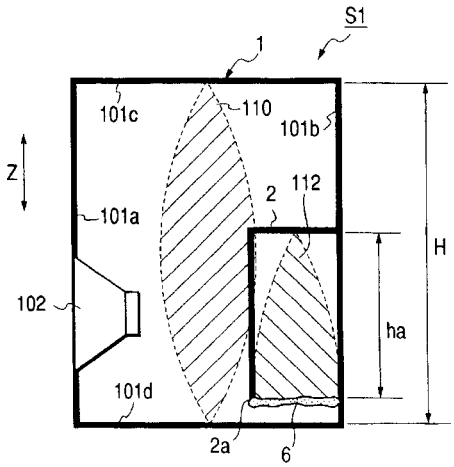
The shell of a laptop computer can be excited by an audio frequency driver and thus also serve as a panel-type loudspeaker. In this variant design, additional hinged or sliding panels can be opened to form a waveguide which directs sound energy toward the user.—GLA

6,324,292

43.38.Ja SPEAKER APPARATUS

Takashi Mitsuhashi and Hiroyuki Hamada, assignors to Pioneer Corporation
27 November 2001 (Class 381/349); filed in Japan 14 October 1998

A prominent resonance in a loudspeaker cabinet can be suppressed by including a relatively small Helmholtz resonator inside the box. However, if the peak is caused by a standing wave **110**, then there will be additional



harmonic resonances as well. By using a quarter-wave pipe **2** as a filter, half of these can be attenuated. Several pipes can be installed to control multiple resonant peaks.—GLA

6,324,294

43.38.Ja PASSENGER VEHICLES INCORPORATING LOUDSPEAKERS COMPRISING PANEL-FORM ACOUSTIC RADIATING ELEMENTS

Henry Azima *et al.*, assignors to New Transducers Limited
27 November 2001 (Class 381/381); filed 2 September 1996

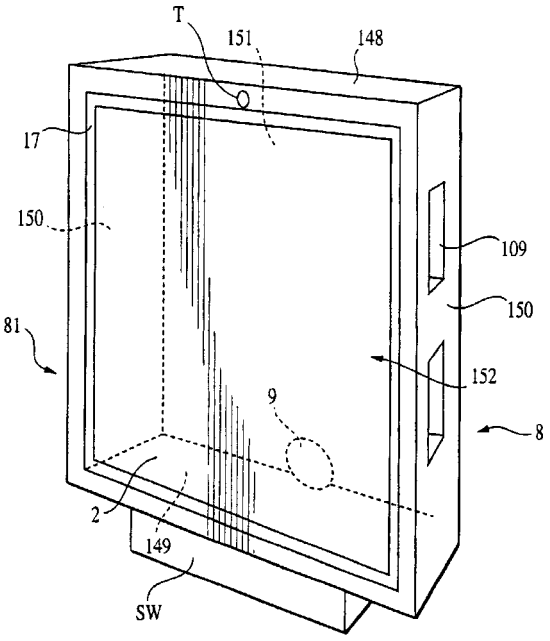
Similar to United States Patent 6,320,967, reviewed above, this one concerns the integration of distributed mode acoustic radiators into an automobile headliner, providing enhanced bass response. There is no word on external versus internal radiation. Other candidate locations for placement of these loudspeakers are head restraints, foot wells, and sun visors.—KPS

6,327,369

43.38.Ja LOUDSPEAKERS COMPRISING PANEL-FORM ACOUSTIC RADIATING ELEMENTS

Henry Azima *et al.*, assignors to New Transducers Limited
4 December 2001 (Class 381/152); filed in the United Kingdom 2 September 1995

A number of years ago this reviewer tried mounting a commercial panel-type loudspeaker on the front of a shallow box. It now appears that the



experiment may have pre-infringed on the patent at hand.—GLA

6,282,155

43.38.Md RECORDING METHOD AND APPARATUS FOR CONTINUOUS PLAYBACK OF FRAGMENTED SIGNALS

Takao Takahashi *et al.*, assignors to Sony Corporation
28 August 2001 (Class 369/32); filed in Japan 19 May 1997

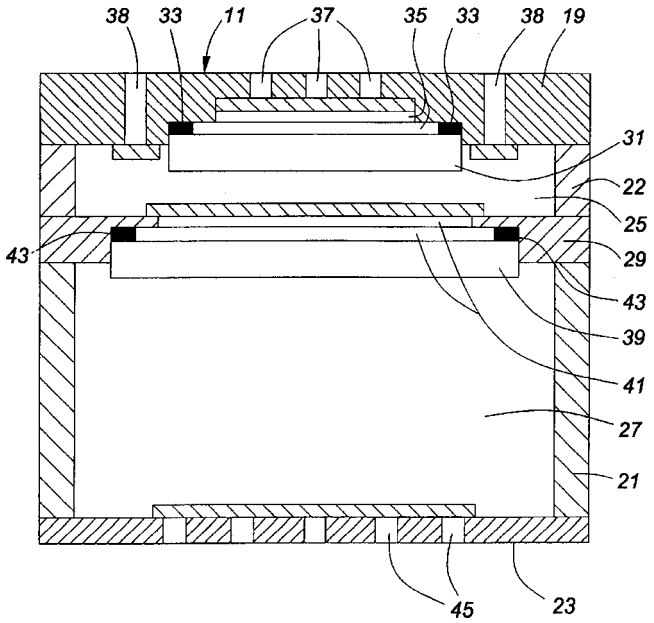
Disks offer large amounts of storage but suffer from rotational latency as well as head seek time. This patent proposes organizing the file system in time ''chunks'' and then reallocating leftover fragments. The patent is seemingly unaware of the large body of work done in operating systems to address real time disk scheduling.—MK

6,324,284

43.38.Si TELEPHONE HANDSET WITH ENHANCED HANDSET/HANDSFREE RECEIVING AND ALERTING AUDIO QUALITY

Larry E. Hawker and Walter Dolezal, assignors to Nortel Networks Limited
27 November 2001 (Class 379/433.02); filed 5 May 1997

This dual-mode transducer includes a conventional telephone receiver **31** which conducts sound to the listener's ear through openings **37**. However, if the ear in question is only partially sealed against the handset, then



it will still receive low-frequency energy from loudspeaker **39** through openings **38**. Conventional hands-free operation is also provided by rotating the handset so that loudspeaker sound from openings **45** predominates.—GLA

6,321,602

43.40.Le CONDITION BASED MONITORING BY VIBRATIONAL ANALYSIS

Mohamed Ben-Romdhane, assignor to Rockwell Science Center, LLC
27 November 2001 (Class 73/660); filed 28 September 1999

This patent pertains to monitoring of machinery and, more specifically, of bearings. Baseline parameters are determined by a microprocessor analysis of vibration spectra, comparing measured peaks to those predicted by a model. The microprocessor estimates a rotation rate and the number of rolling elements in a bearing by attempting to fit the model to the actual spectrum, using various assumptions. Baseline profiles are formed from selected tones and their harmonics. Later acquired data are compared to the baseline characteristics, and deviations from these are used as diagnostic indicators of bearing condition.—EEU

6,323,943

43.40.Yq VIBRATION MEASUREMENT METHOD AND APPARATUS

Tetsuro Maruyama and Akiyoshi Ohno, assignors to Suzuki Motor Corporation
27 November 2001 (Class 356/28.5); filed in Japan 24 September 1998

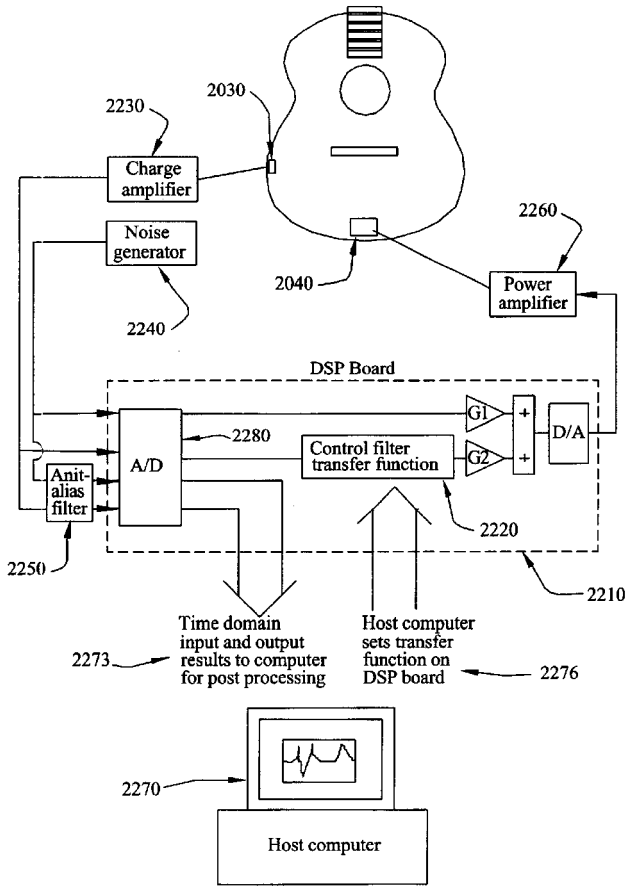
This self-mixing-type laser Doppler system is claimed to be able to measure vibrational displacements that are smaller than half the wavelength of the laser's light. The laser beam impinging on the test object is mixed with the reflected beam to obtain a beat wave. The ratio of the beat wave amplitude at the turning point of the test object to a predetermined reference amplitude is determined and used to calculate the displacement amplitude at the object's turning point.—EEU

6,320,113

43.50.Ki SYSTEM FOR ENHANCING THE SOUND OF AN ACOUSTIC INSTRUMENT

Steven F. Griffin *et al.*, assignors to Georgia Tech Research Corporation
20 November 2001 (Class 84/738); filed 19 July 1995

In the opening of this patent, the following question is posed: "In 1990, the 'Mendelssohn' Stradivarius violin sold at Christie's in London for $1,686,700. A good violin at a typical music store sells for around $2,000. What is it about the Stradivarius that makes it cost almost 1,000 times as much?" This patent first proposes the use of finite element modeling (FEM)



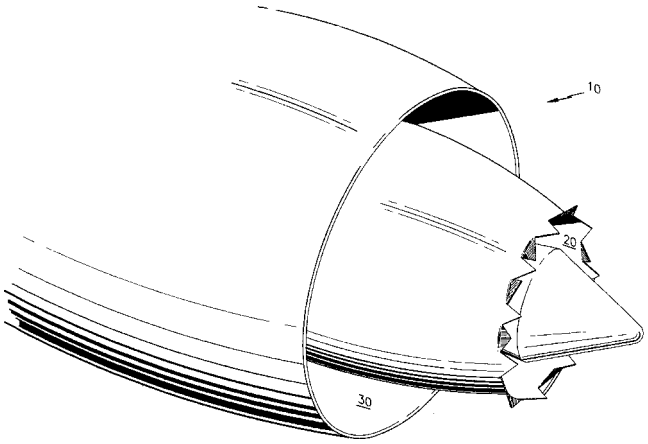
for hollow body stringed instruments. The FEM model is used to obtain a reduced order state space model for the back plate. This is further used in an active noise control inspired feedback system. The patent is well written and informative.—MK

6,314,721

43.50.Nm TABBED NOZZLE FOR JET NOISE SUPPRESSION

Douglas C. Mathews and John K. C. Low, assignors to United Technologies Corporation
13 November 2001 (Class 60/264); filed 4 September 1998

A method to reduce jet noise from turbofan aircraft engines consists of tabs placed at the nozzle exit. They may be used on the nozzle for the core flow (as illustrated in the figure) or for the by-pass flow. The tabs are ar-



ranged in an alternating circumferential pattern; one tab is directed radially inward, the next is a smooth extension of the nozzle, the next is directed radially outward with angular displacement of the tabs is less than 30 degrees. These tabs generate vortices which promote mixing between the core flow and the slower by-pass flow. At take-off power settings, where exhaust velocities are relatively high, this device reduces jet noise by 3 EPNdB with minimal thrust performance penalty.—KPS

6,319,609

43.55.Ev METHOD FOR PRODUCING SOUND ABSORBING UNITS

Berndt Eriksson and Berne Geewe, assignors to Zeta Innovation HB
20 November 2001 (Class 428/407); filed in Sweden 6 March 1997

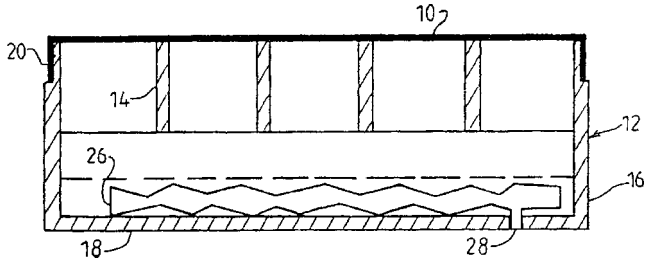
Take old tires, chop them up, mix with suitable binding agent, solidify the mixture in a mold, and then insert suitable devices for holding or lifting the resulting units. Use these units to make a sound barrier wall along a highway or railway.—CJR

6,332,027

43.55.Ev NOISE-ABSORPTION STRUCTURES AND WALLS CONSTITUTED THEREBY

Jean-Claude Guilloud *et al.*, assignors to Bertin & Cie; Societe Nationale d'Etude et de Construction de moteurs d'Aviation—Snecma
18 December 2001 (Class 381/71.1); filed in France 28 June 1996

This patent first makes reference to prior patents relating to lightweight, noise-absorbing structures that comprise a gastight membrane stretched over a frame to form a passive laminar air-flow damper. As an improvement to this approach, the acoustic impedance of the volume



trapped by the membrane is controlled and varied by data processing systems. The resulting noise-absorbing structure would be applicable to the aviation industry and other sources of transportation noise.—CJR

6,334,280

43.55.Ev SOUND ABSORBING CEMENTITIOUS TILE

Etienne Frappart *et al.*, assignors to BPB PLC
1 January 2002 (Class 52/144); filed in the United Kingdom 8 February 1996

This die assembly production process provides a more thorough penetration system for punching through gypsum plasterboard tiles so as to improve the absorptive properties of the tile. The assembly includes a lower die plate with perforations that match those of the punch pins on the upper stripper plate, so the perforations can more easily go entirely through the gypsum board tile.—CJR

6,319,579

43.55.Ti SETTABLE MIXTURE AND A METHOD OF MANUFACTURING A SOUND INSULATING FLOOR CONSTRUCTION

Christian Strandgaard, assignor to SIKA AG, vorm. Kaspar Winkler & Company
20 November 2001 (Class 428/45); filed 3 June 1997

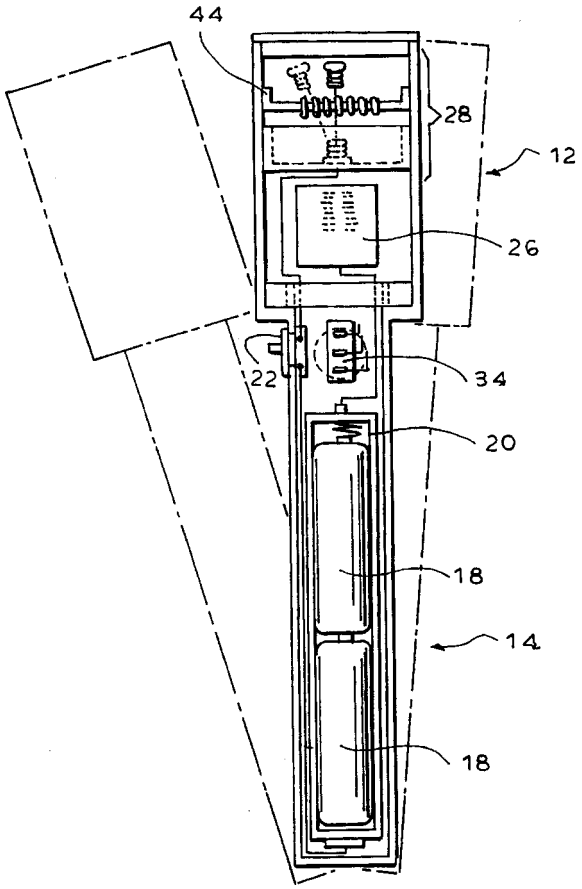
The patent describes a mixture used as a topping over insulation to make a ''floating floor'' with high strength and sound attenuation. Fibers are added to the mixture to reduce the expansion that degrades other similar slabs.—CJR

6,297,438

43.58.Jq TOY MUSICAL DEVICE

Tong Kam Por Paul, Hong Kong, Hong Kong Special Admin. of the People's Republic of China
2 October 2001 (Class 84/600); filed 28 July 2000

Imagine a baton with internal switches such that if you shake it, it



would create a sequence of notes. Twelve tone composers could use the output, too.—MK

6,324,290

43.58.Kr METHOD AND APPARATUS FOR DIAGNOSING SOUND SOURCE AND SOUND VIBRATION SOURCE

Kazutomo Murakami *et al.*, assignors to Bridgestone Corporation
27 November 2001 (Class 381/92); filed in Japan 8 March 1994

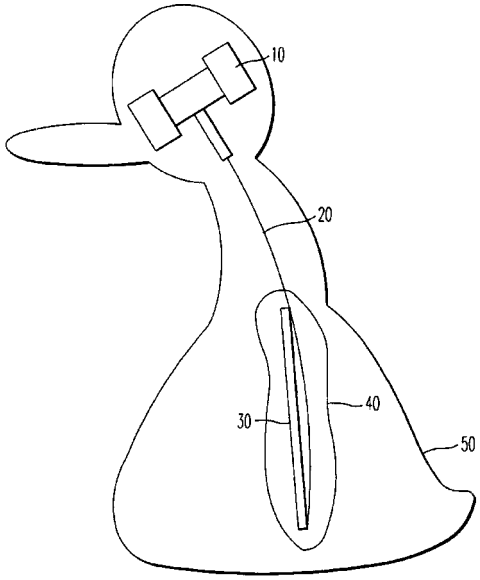
One goal in the design of an automobile is to minimize noise in the passenger compartment as efficiently as possible. An engineer might decide to mount a number of microphones and vibration sensors at appropriate locations and connect their outputs to a multi-channel chart recorder. He might even write a computer program to facilitate subsequent analysis. This much is known prior art. The patent describes an improved procedure involving computer control of digital filters operating in real time.—GLA

6,283,817

43.58.Wc STUFFED TOY

Takashi Kawaguchi, assignor to Kawaguchi Company, Limited
4 September 2001 (Class 446/297); filed in Japan 11 August 1999

This passive rocking toy proposes using the restoring force of the



spring to power an undescribed mechanical powered sound source.—MK

6,318,180

43.58.Wc METHOD AND APPARATUS FOR DETERMINING A FREQUENCY AT WHICH A RESONATOR RESONATES

Francis Alan Humphrey *et al.*, assignors to Lattice Intellectual Property Limited
20 November 2001 (Class 73/622); filed in the United Kingdom 24 June 1998

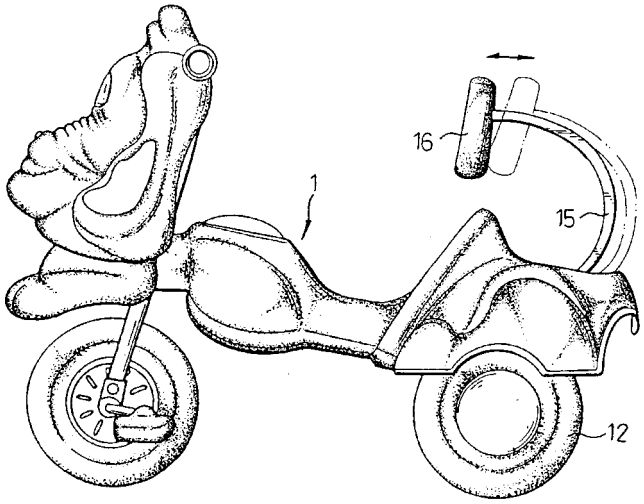
The resonator, such a loudspeaker, is first excited sequentially in a series of relatively wide frequency bands. It is then excited sequentially in a series of narrower bands that cover the wide band which gave the largest response. The process is then repeated. Stepping from band to band can be done quickly, potentially under control of a microprocessor, so that the resonant frequency can be determined rapidly. Frequency scanning is done first in one direction, and then in the other, and the results are averaged in order to reduce the effects of finite build-up times of resonant vibrations during rapid scanning. Additionally, the results from a number of scans are averaged.—EEU

6,319,089

43.58.Wc SOUND GENERATOR FOR A CHILD'S CYCLE

Wen-Tai Cheng and Tainan Hsien, Taiwan, Province of China
20 November 2001 (Class 446/409); filed 5 January 2000

A child's cycle can have a noise maker put in the seat but this may



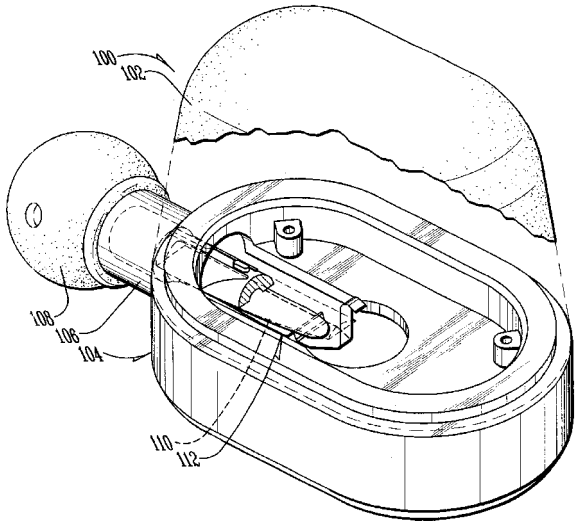
result in discomfort. Instead, the patent proposes putting sound generation below the seat and attached to a lever connected to the seat.—MK

6,328,623

43.58.Wc APPARATUS AND METHOD FOR CALLING GAME

Ron M. Bean, assignor to Hunter's Specialties, Incorporated
11 December 2001 (Class 446/208); filed 7 June 1999

Elk hunters like to entice their prey with synthetic calls. The patent



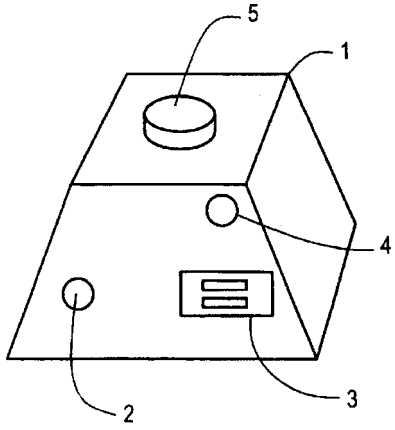
describes a reed coupled to a cavity with a bellows source for fooling large mammals.—MK

6,338,036

43.60.Bf CONFIRMATION NOTIFICATION BY APPARATUS USING AUDIO RECOGNITION AS TO THE ACCEPTABILITY OF AN INPUT SOUND

Yasunaga Miyazawa, assignor to Seiko Epson Corporation
8 January 2002 (Class 704/246); filed in Japan 3 September 1998

The aim of this device is to indicate whether a speaker has spoken an appropriate sound. The input consists of a group of words. The digitized sound data is analyzed with respect to sound power and characteristic data. The word sequence is segmented on the basis of sound power and a determination is made as to whether the pattern of sound power input and word durations is recognized as a speech sound. A speech recognition-processing phase follows, which then outputs the information that the recognized object sound is appropriate.—DRR



6,330,339

43.66.Ts HEARING AID

Ryuuichi Ishige and Reishi Kondo, assignors to NEC Corporation
11 December 2001 (Class 381/312); filed in Japan 27 December 1995

The operational mode of a hearing aid is varied as a function of the activity of the wearer and the time of day. When the wearer is inactive, e.g., during sleep, power supply voltage, gain, and filtering are reduced by using a simple signal processing algorithm with reduced clock frequency, resulting in lower power consumption.—DAP

6,334,441

43.70.Dn PHONATION VALVE FOR BREATHING TUBE

John Zowtiak *et al.*, assignors to Mallinckrodt Medical, Incorporated
1 January 2002 (Class 128/207.16); filed 23 November 1998

Previously patented breathing tubes have provided a valve to close during exhalation, forcing the breath to exit through the glottis, allowing the patient to speak. This device also provides a damping mechanism to reduce the sound of ''diaphragm harmonics'' during speech.—DLR

6,334,776

43.70.Dn METHOD AND APPARATUS FOR TRAINING OF AUDITORY/VISUAL DISCRIMINATION USING TARGET AND DISTRACTOR PHONEMES/GRAPHEMES

William M. Jenkins *et al.*, assignors to Scientific Learning Corporation
1 January 2002 (Class 434/169); filed 17 December 1997

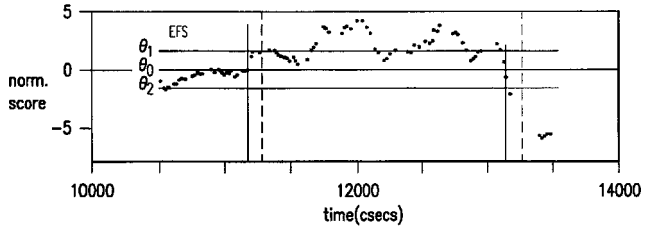
This is another in the assignee's long series of patents on teaching a patient with impaired speech perception to overcome that impairment. Here, a target sound is enhanced for easier detection and is then surrounded by distractor sounds. As the user correctly identifies the target sound, the degree of enhancement is progressively reduced to the point of no enhancement.—DLR

6,317,710

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

Qian Huang *et al.*, assignors to AT&T Corporation
13 November 2001 (Class 704/246); filed 13 August 1998

This system would allow a user to request the playback of multimedia material from a large (unspecified) database based on the automatic identification of the voice of a particular speaker. Several well-known speaker identification techniques, such as Gaussian mixture models, are suggested. A



normalized log likelihood identification score is compared to threshold levels to identify the desired content blocks. Multiple threshold levels allow the detection of some types of boundaries within the content material.—DLR

6,330,536

43.72.Fx METHOD AND APPARATUS FOR SPEAKER IDENTIFICATION USING MIXTURE DISCRIMINANT ANALYSIS TO DEVELOP SPEAKER MODELS

Sarangarajan Parthasarathy and Aaron E. Rosenberg, assignors to AT&T Corporation
11 December 2001 (Class 704/249); filed 25 November 1997

This speaker identification system uses both speaker-dependent (SD) and speaker-independent (SI) processing systems. Upon registration of a new user, a SI identification program performs phonetic segment analysis of three spoken instances of the password. The results are compared with all other registered passwords and a Gaussian mixture model is created for the new user using k-means clustering and expectation-maximization. During use, in SD mode, a discriminant analysis is performed to compute likelihood scores, which are combined to determine the speaker's identity.—DLR

6,307,140

43.72.Ja MUSIC APPARATUS WITH PITCH SHIFT OF INPUT VOICE DEPENDENTLY ON TIMBRE CHANGE

Kazuhide Iwamoto, assignor to Yamaha Corporation
23 October 2001 (Class 84/622); filed in Japan 30 June 1999

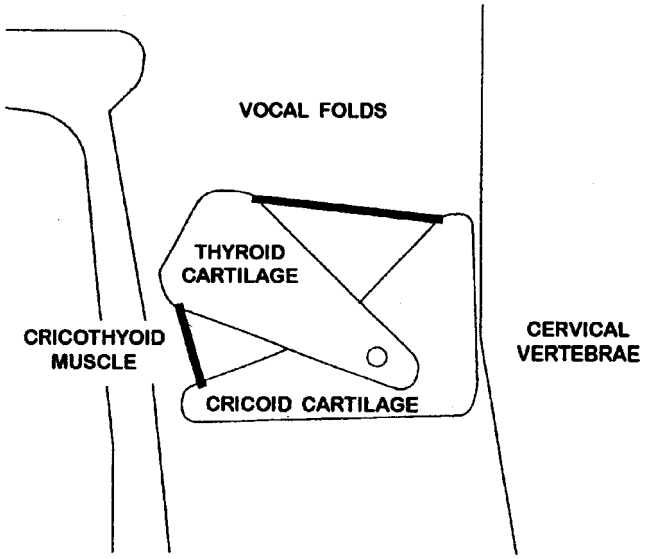
An undisclosed method of formant shifting is used for gender bending in voice harmony synthesis. Additional effects such as vibrato can be added on each formant. Stored multivoice pitch shifts can create barbershop quartets or a falsetto duet... or more. Chorusing is yet another option.—MK

6,317,713

43.72.Ja SPEECH SYNTHESIS BASED ON CRICOTHYROID AND CRICOID MODELING

Seiichi Tenpaku, assignor to Arcadia, Incorporated
13 November 2001 (Class 704/261); filed in Japan 25 March 1996

This speech synthesizer uses a model of the glottis to compute basic parameters which determine the glottal excitation pulse waveform. The glottal frequency is then controlled according to two parameters, a declining



MODEL OF LARYNX

phrase contour and a momentary accent component, which marks specific words or syllables.—DLR

6,332,121

43.72.Ja SPEECH SYNTHESIS METHOD

Takehiko Kagoshima and Masami Akamine, assignors to Kabushiki Kaisha Toshiba
18 December 2001 (Class 704/262); filed in Japan 4 December 1995

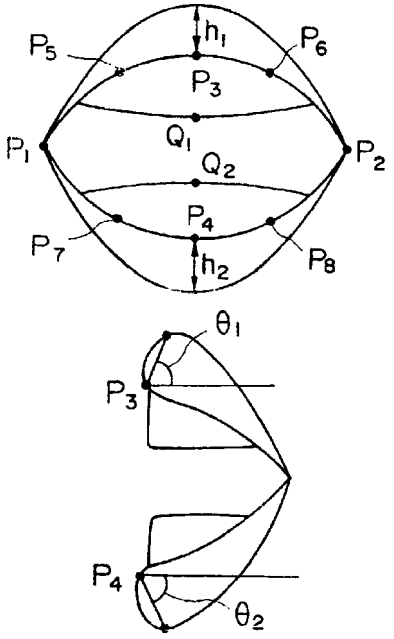
This speech synthesizer includes storage for a substantial amount of speech for use as reference material during synthesis. Analysis procedures, based primarily on linear prediction, offer the extraction of those characteristics from the stored speech which most likely contribute to its naturalness and provide various ways to reuse this information during the generation of output speech from text or phonetic inputs.—DLR

6,332,123

43.72.Ja MOUTH SHAPE SYNTHESIZING

Masahide Kaneko *et al.*, assignors to Kokusai Denshin Denwa Kabushiki Kaisha
18 December 2001 (Class 704/276); filed in Japan 8 March 1989

This audio/video speech synthesizer accepts a text input and produces as output both speech and a video display of the face. The input is separated into syllables for lip shape processing. Basic lip shape parameters are stored



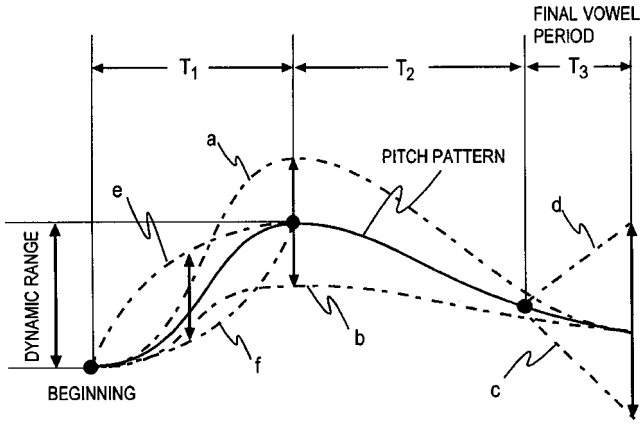
in a table accessed by a syllable code. These parameters are then used as the basis from which to ''flesh out'' a realistic view of a talking face.—DLR

6,334,106

43.72.Ja METHOD FOR EDITING NON-VERBAL INFORMATION BY ADDING MENTAL STATE INFORMATION TO A SPEECH MESSAGE

Osamu Mizuno and Shinya Nakajima, assignors to Nippon Telegraph and Telephone Corporation
25 December 2001 (Class 704/260); filed in Japan 21 May 1997

This speech synthesis system accepts text which has been annotated with keywords indicating the emotional or affective state by which the speech is to be generated. The top level, or ''S'' layer, accepts keywords such as ''doubtful,'' ''angry,'' ''questioning,'' or ''encouraging.'' Lower lay-



ers interpret and realize the corresponding states by controlling speech properties, including pitch, segment duration, speaking rate, dynamic range, vibrato, tremor, etc.—DLR

6,317,717

43.72.Ne VOICE ACTIVATED LIQUID MANAGEMENT SYSTEM

Kenneth R. Lindsey, Louisville, Kentucky *et al.*
13 November 2001 (Class 704/275); filed 25 February 1999

Using this device, you can give commands to fill the tub to get your bath ready. Separate solenoids control hot and cold water according to a temperature sensor and a tub overflow sensor. You can also turn on the shower or flush the toilet.—DLR

6,324,513

43.72.Ne SPOKEN DIALOG SYSTEM CAPABLE OF PERFORMING NATURAL INTERACTIVE ACCESS

Akito Nagai *et al.*, assignors to Mitsubishi Denki Kabushiki Kaisha
27 November 2001 (Class 704/275); filed in Japan 18 June 1999

This interactive dialog system is designed to detect any of a moderate number of prespecified keywords spoken during normal, fluent speech. In this case, the keywords represent geographical places within a telephone calling area. Considerable effort is devoted to the problem of distinguishing the intended keywords from common conversational words and proper names which might come up during the same conversation.—DLR

6,330,539

43.72.Ne DIALOG INTERFACE SYSTEM

Kuniharu Takayama *et al.*, assignors to Fujitsu Limited
11 December 2001 (Class 704/275); filed in Japan 5 February 1998

This patent presumably describes a control strategy which would hopefully be successful in managing a general-purpose, unconstrained speech interface system, intended as an assistant in taking care of and living in a house. There is little mention of hardware; accurate recognition and synthesis between speech and phonetic sequences are assumed. Carrying on in this tradition, the sum total of the detail given on how input processing proceeds is that ''...the input sequence of characters is further converted to an input semantic representation.''—DLR

6,332,120

43.72.Ne BROADCAST SPEECH RECOGNITION SYSTEM FOR KEYWORD MONITORING

Robert L. Warren, assignor to Solana Technology Development Corporation
18 December 2001 (Class 704/235); filed 20 April 1999

The application of an unspecified speech recognition system is described which would monitor a radio broadcast for the occurrence of specific words. In an example application, a manufacturer would monitor nationwide news broadcasts for any mention of their product. This would presumably allow a quicker response to reports such as product tampering or contamination.—DLR

6,332,122

43.72.Ne TRANSCRIPTION SYSTEM FOR MULTIPLE SPEAKERS, USING AND ESTABLISHING IDENTIFICATION

Kerry A. Ortega *et al.*, assignors to International Business Machines Corporation
18 December 2001 (Class 704/270); filed 23 June 1999

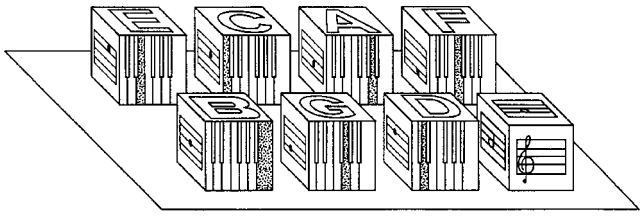
This dictation and transcription system is designed for use in an office environment where multiple users may simultaneously enter spoken material to be transcribed. The brief description is concerned with user identification protocols and voice-controlled navigation through the details of data entry.—DLR

6,271,453

43.75.Tv MUSICAL BLOCKS AND CLOCKS

L. Leonard Hacker, assignor to L. Leonard Hacker
7 August 2001 (Class 84/476); filed 21 May 1997

Blocks are truly a child's best friend. Why not add a reed or electrical tone generator and create musical blocks? The inventor claims it can teach



''music reading, instrument playing, tempo, staccato, note duration, rhythm, harmony, chords...and music writing.''—MK

6,335,691

43.75.Tv MUSICAL BOTTLE

Darrell Bird, assignor to Darrell Bird
1 January 2002 (Class 340/691.1); filed 9 August 2000

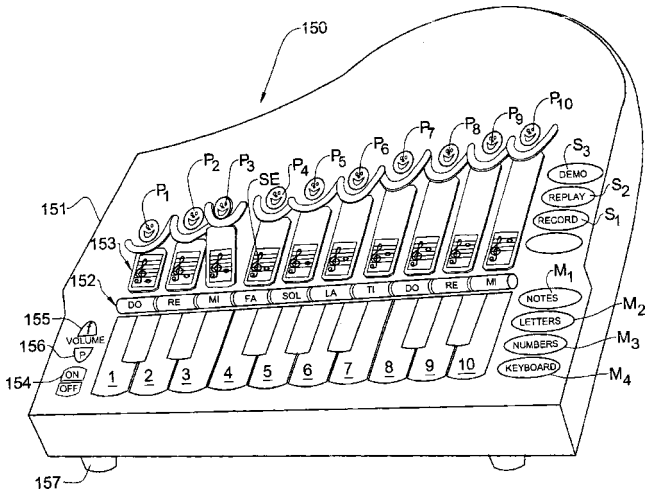
Bars and pubs are places to drink and talk. So, why not create a talking alcoholic beverage bottle? The patent proposes a sensor activated musical or talking bottle (with the transducer on the bottom).—MK

6,337,434

43.75.Tv MUSIC TEACHING INSTRUMENT

Dorly Oren-Chazon, Tel Aviv, Israel
8 January 2002 (Class 84/478); filed 14 May 1998

The patent proposes a teaching machine for the well-known musical alphabet Solfeg. The machine is a computer-controlled synthesizer with



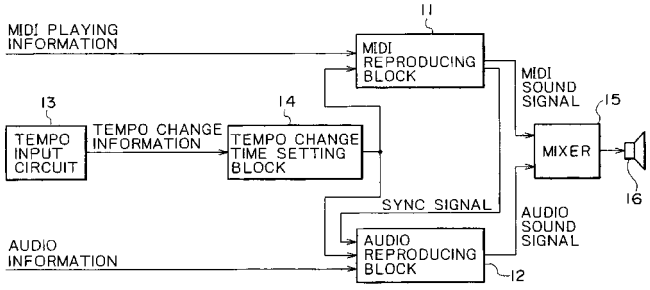
puppets, colors, lights, and other metaphors.—MK

6,281,424

43.75.Wx INFORMATION PROCESSING APPARATUS AND METHOD FOR REPRODUCING AN OUTPUT AUDIO SIGNAL FROM MIDI MUSIC PLAYING INFORMATION AND AUDIO INFORMATION

Takashi Koike *et al.*, assignors to Sony Corporation
28 August 2001 (Class 84/636); filed in Japan 15 December 1998

In accompaniment, the accompanying voice must try and match the speed of the soloist. Given a MIDI stream for the automatic accompaniment, the question is how to change the MIDI stream with changing tempo. The



patent describes a method for synchronizing the MIDI stream and audio stream with the tempo. Unfortunately, the patent gives no hints or details about how to measure beat change.—MK

6,288,320

43.75.Wx ELECTRIC MUSICAL INSTRUMENT

Kazuo Murakami and Hideo Suzuki, assignors to Yahama Corporation
11 September 2001 (Class 84/622); filed in Japan 26 February 1999

The convolution of the string with the body in a stringed instrument means it's difficult to derive the original string performance parameters. One solution to this problem is to place a sensor in the bow as well as the body. You can also apply an inverse filter to invert the body (numerical problems notwithstanding). The patent is well written and informative.—MK

6,298,322

43.75.Wx ENCODING AND SYNTHESIS OF TONAL AUDIO SIGNALS USING DOMINANT SINUSOIDS AND A VECTOR-QUANTIZED RESIDUAL TONAL SIGNAL

Eric Lindemann, assignor to Eric Lindemann
2 October 2001 (Class 704/222); filed 6 May 1999

In sinusoidal coding, the residual can also be coded for better fidelity. However, the residual is noiselike, which presents coding difficulties. The patent proposes using vector quantization (VQ) to code the residual. Furthermore, if the sound is known to be harmonic, then additional coding is possible. The patent includes detailed algorithmic flowcharts.—MK

6,313,389

43.75.Wx AUTOMATIC ACCOMPANIMENT APPARATUS ADJUSTING SOUND LEVELS FOR EACH PART OF A PLURALITY OF PATTERNS WHEN ONE PATTERN IS SWITCHED TO ANOTHER PATTERN

Yoshihisa Adachi, assignor to Kabushiki Kaisha Kawai Gakki Seisakusho
6 November 2001 (Class 84/634); filed in Japan 30 April 1999

In simplistic automatic accompaniment, the computer has fixed patterns (e.g., ''introduction,'' ''fill-in,'' ''ending''). If one instrumental part is muted, then when the pattern switches, the muting may be lost resulting in a jarring change ''disturbing a natural play.'' This strictly algorithmic patent proposes a very simple method to account for these changes.—MK

6,316,710

43.75.Wx MUSICAL SYNTHESIZER CAPABLE OF EXPRESSIVE PHRASING

Eric Lindemann, assignor to Eric Lindemann
13 November 2001 (Class 84/609); filed 27 September 1999

While sampling instruments offer realistic sounds, they are not expressive. The patent describes a state-machine-based approach to expressive synthesis from stored samples. Like the author's previous patent, United States Patent 6,298,322, reviewed above, it is clear and also gives detailed flowcharts.—MK

6,316,711

43.75.Wx MUSICAL SOUND SIGNAL GENERATION APPARATUS

Eiji Matsuda *et al.*, assignors to Kabushiki Kaisha Kawai Gakki Seisakusho
13 November 2001 (Class 84/615); filed in Japan 16 February 1999

Complex scores or sounds demand many digital oscillators. What should happen when free oscillators are exhausted? The patent proposes (1) changing from stereo to mono and (2) using a random selection of older voices (a cheap form of LRU).—MK

6,327,367

43.75.Wx SOUND EFFECTS CONTROLLER

G. Scott Vercoe, West Roxbury, Massachusetts and Alexandre Burton, Montreal, Quebec, Canada
4 December 2001 (Class 381/61); filed 14 May 1999

Computer gaming has produced legions of joystick practicioners. The patent sketches a method of converting the joystick X, Y, and Z outputs to synthesizer parameters. No details are given on the best way to derive this map.—MK

6,336,367

43.75.Yy VIBRATION TRANSDUCER UNIT

Heikki Räisänen, assignor to B-Band Oy
8 January 2002 (Class 73/632); filed in Finland 29 January 1998

In these units, as may be mounted in the saddles or bridges of acoustic guitars or other string instruments, the sensor and preamplifier are integrated into a small package. An electromechanical film serves as the sensing element. It may be supported on the circuit board that accommodates the electronic components and that also may form part of the unit's housing.—EEU

Relating speech intelligibility to useful-to-detrimental sound ratios (L)

J. S. Bradley
Institute for Research in Construction, National Research Council, Ottawa K1A 0R6, Canada

S. R. Bistafa
Department of Mechanical Engineering, Polytechnic School, University of São Paulo, São Paulo, Brazil

(Received 3 January 2002; accepted for publication 8 April 2002)

This letter corrects typographical errors in two previously published regression equations relating measured speech intelligibility scores and useful-to-detrimental sound ratios. The implications of one of these corrections on a more recent paper are shown to be small. A new and more generally applicable relationship is presented that was derived from a larger data set. This new relationship is presented as a better means of estimating expected speech intelligibility from useful-to-detrimental sound ratios in a wide range of rooms.
[DOI: 10.1121/1.1481508]

PACS numbers: 43.55.Hy, 43.71.Gv [MK]

I. INTRODUCTION

Two earlier publications related speech intelligibility test scores in rooms to measured useful-to-detrimental sound ratios. The first study[1] considered a range of larger rooms. The second study[2] was based on measured values in school classrooms. One of the relationships from the classroom study was used in a more recent publication[3] that compared the implications of various speech intelligibility metrics when applied to classrooms. It has been brought to our attention that there were typographical errors in the original documents[1,2] that have also influenced the analyses of the more recent publication.[3]

More generally applicable relationships between speech intelligibility scores and useful-to-detrimental sound ratios can be derived by combining the data from both of the earlier studies. This was done previously to a limited extent but not widely publicized.[4]

This letter therefore has three objectives. The first is to correct the errors in the published regression coefficients of the equations relating speech intelligibility scores and useful-to-detrimental sound ratios. The second is to consider the implications of the corrected relationship on the comparisons with other speech metrics. The third objective is to make available a more broadly applicable relationship based on the results of both earlier measurement studies.

II. CORRECTIONS TO PREVIOUS ERRORS

The initial measurement study included a regression equation [Eq. (9) in Ref. 1] relating measured speech intelligibility scores (SI) using a Fairbanks rhyme test to measured U_{95} values. The U_{95} values were obtained according to the original useful-to-detrimental sound ratio concept proposed by Lochner and Burger.[5] Useful-to-detrimental sound ratios are the ratio of the sum of the direct and early reflections of the speech sounds to the combination of the later-arriving speech sounds and ambient noise. They have been shown to be good predictors of speech intelligibility scores in rooms because they combine the effects of signal-to-noise ratio and room acoustics. Lochner and Burger's original formulation included a somewhat complicated weighting of the importance of each early arriving reflection in the calculation of the useful component of the ratio. The obtained relationship [Eq. (9) in Ref. 1] should have been given as

$$\mathrm{SI} = +0.7348\mathrm{U}_{95} - 0.09943\mathrm{U}_{95}^2 - 0.0005457\mathrm{U}_{95}^3 + 97.39. \quad (1)$$

The final constant was incorrect as previously published.

Equation (2) in Ref. 2 also included a typographical error. It related measured speech intelligibility scores (SI) to measured U_{50} values. U_{50} is a useful-to-detrimental ratio where the early energy is an unweighted sum of the direct sound and early reflections arriving within 50 ms after the direct sound. This equation should have been printed as follows:

$$\mathrm{SI} = +1.027\mathrm{U}_{50} - 0.0838\mathrm{U}_{50}^2 + 99.42. \quad (2)$$

A zero was missing from the coefficient of the squared term in the original publication.

III. IMPLICATIONS OF THE CORRECTIONS

The incorrect version of Eq. (2) above was used in the recent comparison of various speech metrics and was repeated (incorrectly) as Eq. (27) in Ref. 3. When this is corrected, Fig. 6 from Ref. 3 is changed, along with some small details derived from this form of graph for various room volumes that were included in Table II (of Ref. 3). The re-calculated version of the former Fig. 6 is included here as Fig. 1. It shows calculated speech intelligibility scores (SI) versus reverberation time (T) for various combinations of $L_n - L_{splm}$. Here L_n is the ambient noise level and L_{splm} is the speech level at 1 m in a free field.

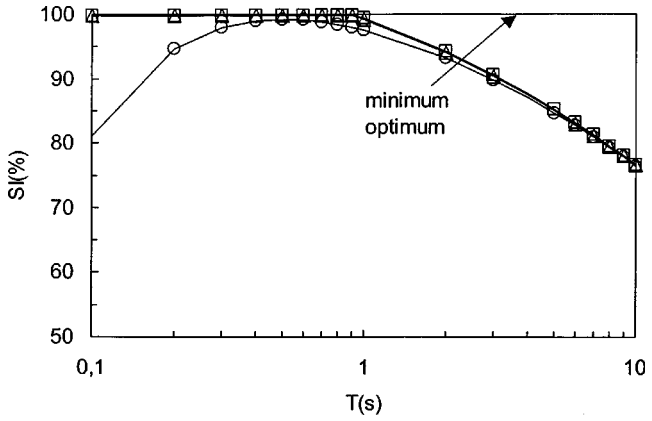


FIG. 1. Speech intelligibility given by the Fairbank's rhyme test, calculated using Eq. (2) with U_{50} obtained from Eq. (12) (of Ref. 3) as a function of the reverberation time T, for a 300-m^3 classroom and for $L_n - L_{splm}$ equal to -10 dB (○), -20 dB (△), and -30 dB (□). Indicated in this figure is the "minimum optimum" value of 100% speech intelligibility recommended by Bradley (Ref. 4).

The new figure shows that predicted SI values decrease less rapidly than in the original figure for very small or very large reverberation times. Similar to the original, the figure shows that 100% speech intelligibility scores are found over a range of reverberation times from about 0.1 to 0.9 s for $L_n - L_{splm}$ values of -20 and -30 dB. Table II (of Ref. 3) indicated that, according to U_{50}, 100% speech intelligibility would occur over a range of reverberation times from 0.1 to 1.0 s for most of the combinations of $L_n - L_{splm}$ and room volumes. The corrected calculations indicate that these would change to a range of 0.1 to approximately 0.9 s. Because this is a very small change, a corrected version of the previous Table II is not included here and the discussion and conclusions that were previously made, based on Table II of Ref. 3, would not change.

IV. MORE BROADLY APPLICABLE RELATIONSHIP

A more generally applicable regression equation relating speech intelligibility and useful-to-detrimental ratios can be derived by combining the data from both of the earlier measurement studies.[1,2] This provides a total of 320 sets of average speech intelligibility scores and measured U_{50} values. The resulting regression equation is

$$\mathrm{SI} = +0.861\mathrm{U}_{50} - 0.0863\mathrm{U}_{50}^2 + 98.24. \quad (3)$$

The associated $R^2 = 0.737$ with $p < 0.0001$.

This new equation is compared with Eq. (2) and the combined original measurement data in Fig. 2. The two equations are seen to be very similar. This is partly because both data sets extended over the range of 1-kHz U_{50} values from -20 to $+2$ dB. Because Eq. (3) is based on a larger number of measurements and a larger range of room sizes, it is preferred for estimating speech intelligibility from U_{50} values.

Figure 2 shows that Eq. (3) indicates that 100% speech intelligibility is reached for a 1-kHz U_{50} value of approximately $+2$ dB. Although Eq. (3) is a good fit to the mean trend, one should use it with some caution. First, the standard deviation about the regression line is $\pm 8.7\%$, indicating some uncertainty to the estimate of the mean trend. Second, because intelligibility scores cannot exceed 100%, this regression line should probably be a little higher for U_{50} values approximately $\geqslant 0$ dB. Thus it is a conservative estimate of when the mean trend reaches 100% speech intelligibility. For these results, 100% speech intelligibility was concluded to indicate near-optimum conditions for speech communication. A goal of 100% speech intelligibility may seem unnecessarily high, but in practice it represents the result of subjects concentrating very hard to do well on a very simple rhyme test. It would not correspond to 100% intelligibility with

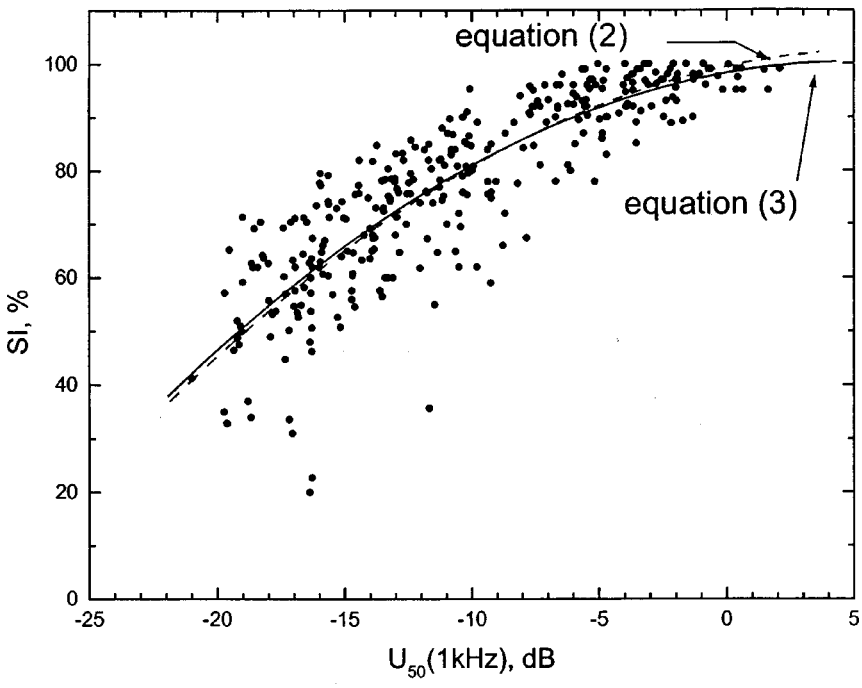


FIG. 2. Comparison of regression equations (2) and (3) and the combined data sets from the two earlier studies (Refs. 1 and 2).

more difficult material or in more relaxed listening conditions that would be more representative of everyday situations.

V. CONCLUSIONS

Corrections to two previously published regression equations have been presented. The changes implied by one of these to the analyses of a more recent publication have been demonstrated to be relatively small.

A new regression equation is presented that is thought to be a better indication of the general relationship between 1-kHz U_{50} values and mean speech intelligibility scores. It indicates that a 1-kHz U_{50} value of $+2$ dB is a reasonable goal for conditions that would permit very good speech communication in rooms.

ACKNOWLEDGMENT

The authors would like to thank Chris Day of Marshal Day Associates, Auckland, New Zealand for pointing out the principal error that is corrected in this letter.

[1] J. S. Bradley, ''Predictors of speech intelligibility in rooms,'' J. Acoust. Soc. Am. **80**, 837–845 (1986).

[2] J. S. Bradley, ''Speech intelligibility studies in classrooms,'' J. Acoust. Soc. Am. **80**, 846–854 (1986).

[3] S. R. Bistafa and J. S. Bradley, ''Reverberation time and maximum background noise level for classrooms from a comparative study of speech intelligibility metrics,'' J. Acoust. Soc. Am. **107**, 861–875 (2000).

[4] J. S. Bradley, ''Uniform derivation of optimum conditions for speech in rooms, Report BRN-239,'' National Research Council, Ottawa, Canada, November 1985.

[5] J. P. A. Lochner and J. F. Burger, ''The influence of reflections on auditorium acoustics,'' J. Sound Vib. **1**, 426–454 (1964).

Measures of auditory-visual integration for speech understanding: A theoretical perspective (L)

Ken W. Grant[a)]
Walter Reed Army Medical Center, Army Audiology and Speech Center, Washington, DC 20307-5001

(Received 1 October 2001; accepted for publication 29 January 2002)

Recent studies of auditory-visual integration have reached diametrically opposed conclusions as to whether individuals differ in their ability to integrate auditory and visual speech cues. A study by Massaro and Cohen [J. Acoust. Soc. Am. **108**(2), 784–789 (2000)] reported that individuals are essentially equivalent in their ability to integrate auditory and visual speech information, whereas a study by Grant and Seitz [J. Acoust. Soc. Am. **104**(4), 2438–2450 (1998)] reported substantial variability across subjects in auditory-visual integration for both sentences and nonsense syllables. This letter discusses issues related to the measurement of auditory-visual integration and modeling efforts employed to separate *information extraction* from *information processing.*
[DOI: 10.1121/1.1482076]

PACS numbers: 43.71.An, 43.71.Ma [CWT]

I. INTRODUCTION

Speech recognition can be severely degraded by noise, reverberation, hearing loss, or combinations of these elements. As demonstrated over the last several decades, auditory-visual (AV) speech recognition by human receivers has been repeatedly shown to be incredibly robust and greatly resistant to environmental and biological sources of signal distortion. One explanation given for the superiority of AV speech recognition over hearing alone (A) is that speechreading (V) provides relatively undistorted cues for place of articulation (e.g., [p] vs [t] vs [k]) which when transmitted auditorily are extremely vulnerable to noise, reverberation, and hearing loss (Summerfield, 1987).

In spite of the robust nature of AV speech recognition, some individuals nevertheless still demonstrate significant problems understanding speech in face-to-face communication settings. Past studies have tried to address this problem by characterizing each individual receiver in terms of several factors believed to be important for speech understanding (Grant *et al.*, 1998; Grant and Seitz, 1998; Grant and Seitz, 2000). Various factors that have been examined included nonsense syllable and sentence recognition by ear and eye, recognition of prosodic speech cues, use of semantic and morpho-syntactic context in sentence recognition, processing speed for words presented in A, V, and AV modalities, and AV integration. Poor performance on one or more of these factors could lead to significant problems understanding speech presented audiovisually (Grant *et al.*, 1998).

The term ''auditory-visual integration'' is used to denote the *processes* employed by individual receivers to combine the information extracted from A and V sources. Integration of A and V cues is thus distinct from the ability to *extract* A and V cues and higher-order language processing of the information received by the two senses (Massaro, 1998). An important question addressed in a recent study by Grant and Seitz (1998) was whether individual receivers integrate A and V cues with varying degrees of efficiency, thereby offering a possible explanation for differences observed across subjects on AV speech recognition tests when more obvious factors such as hearing loss, visual acuity, vocabulary, and language competence had been taken into account. A central assumption of the Grant and Seitz (1998) study was that AV speech integration is a measurable skill, independent of A and V peripheral cue extraction, that subjects use whenever A and V sources of information are available. Grant and Seitz (1998) employed a number of methods using both congruent and discrepant AV speech stimuli, composed of nonsense syllables and meaningful sentences to obtain estimates of integration efficiency (IE). One goal of using varied methodologies and levels of speech materials to estimate IE across listeners was to establish the construct validity of IE as an independent process operating for both congruent and discrepant speech segments and sentences. A second goal was to compare IE within individual subjects when presented with connected speech and isolated syllables, thus examining potential differences in integration processes as a function of speech-processing demands. Results showed that individual hearing-impaired (HI) subjects differed with respect to IE on a variety of measures, and that these individual differences were useful in explaining a significant amount of the variance observed across individuals in AV speech recognition.

Recently, Massaro and Cohen (2000) applied the fuzzy logical model of perception (FLMP) to data sets obtained by Grant and Seitz (1998). Whereas the original analysis of these data using Braida's (1991) prelabeling model of integration (PRE) showed significant differences among HI subjects in IE, Massaro and Cohen concluded that there were no significant signs of inefficient integrations across subjects. The purpose of this letter is to discuss more fully the various issues and assumptions surrounding IE measures and the use of model predictions as a method of measuring IE in individual subjects. Although Massaro and Cohen (2000) make a number of critical statements concerning the methodology and conclusions reported in Grant and Seitz (1998), it is felt that a review of the basic issues and assumptions underlying AV integration would be more useful to the general reader than specific rebuttals. In addition, because the issues raised by Massaro and Cohen only apply to nonsense syllable recognition, comments will be mostly restricted to this domain, although the issues raised by Grant and Seitz regarding the

a)Electronic mail: grant@tidalwave.net

instability of the FLMP and PRE to address auditory-visual integration for sentences are obviously important.

II. WHAT DO WE MEAN BY AUDITORY-VISUAL INTEGRATION EFFICIENCY?

In typical auditory-visual speech recognition tests, subjects are presented with audio materials (syllables, words, phrases, or sentences) that can be either nonsensical or meaningful, along with video of a speaker's face acquired at the same time as the audio materials. The subject gives a response which is checked for accuracy, and when appropriate, for specific patterns of errors. For meaningful speech materials, there are many perceptual stages involved during the encoding and decoding of speech. Grant *et al.* (1998) discussed a framework made up of sensory, linguistic, and higher-order cognitive stages that are likely to impact on an individual's speech recognition performance. As described by Braida (1991), IE may be treated as separate from the extraction of sensory information, and, as a result, it is possible to have efficient integration and still perform very poorly on speech recognition tasks. This would occur if the visual and/or auditory information is greatly impoverished (e.g., deaf subjects with average speechreading skills) or if the information stemming from the two modalities is highly redundant (Grant *et al.*, 1998). Conversely, it is possible, at least theoretically, to extract sufficient auditory and visual sensory information but still do poorly on AV speech recognition tests because of poor integration skills. The study by Grant and Seitz (1998) examined this possibility in a group of hearing-impaired adults who were evaluated on a variety of speech recognition measures. From each of these tests, a different measure of IE was derived. Grant and Seitz concluded that individual hearing-impaired subjects differed significantly with respect to their ability to integrate auditory and visual speech cues, and therefore derived different amounts of speech recognition benefit from the addition of speechreading to audition. This latter conclusion was based on an examination of the variability across subjects on each of the various IE measures and from results of stepwise multiple regression analyses used to predict the relative gain in speech recognition performance from a set of predictor variables that included hearing status, visual and auditory acuity, age, and measures of IE. In each regression analysis conducted, the final subset model included a measure of IE along with other predictor variables (indicating that inclusion of the IE measure helped explain a significantly greater amount of variability in speech recognition performance across subjects than would have been obtained without the IE measure). In addition, individual IE measures were shown to account for roughly 20%–30% of the variance observed in auditory-visual benefit. An interesting result from these analyses was that IE measures derived from nonsense syllable tests were useful in explaining AV consonant benefit, but not AV sentence benefit. Conversely, IE measures derived from sentence tests were useful in explaining AV sentence benefit, but not AV consonant benefit. This result suggests that measures of IE are dependent on the speech materials used (i.e., nonsense syllables versus meaningful sentences).

III. HOW CAN WE MEASURE INTEGRATION EFFICIENCY?

Several measures have been proposed for estimating auditory-visual integration efficiency across individual subjects. One lies in the ability to perceive perceptual fusions when presented with synchronized, but conflicting, auditory and visual speech information as in the McGurk effect (e.g., McGurk and MacDonald, 1976). For example, when presented with audio /ba/ and visual /ga/, how often do subjects perceive /da/ or /ða/? A second proposed measure of integration efficiency lies in the ability to benefit from visual speech cues when congruent audio and visual speech signals are presented asynchronously. Another approach uses model predictions of AV consonant recognition performance to estimate a subject's integration efficiency. In this case, auditory-visual speech recognition scores are predicted from auditory-alone and visual-alone recognition scores (e.g., Blamey *et al.*, 1989; Massaro, 1987, 1998; Braida, 1991, 1996). Integration efficiency is defined as the deviation between predicted and obtained scores or some goodness-of-fit measure such as a chi-square statistic or the root-mean-squared deviation (rmsd). This is an especially useful way of estimating integration efficiency if the models are optimal processing models and predictions are based on the best auditory-visual score one might be expected to obtain given the *information* available via the auditory and visual modalities.

The standard form of the FLMP proposed by Massaro (1987, 1998) and used by Massaro and Cohen (2000) in their analyses of the Grant and Seitz (1998) data assumes an integration process which is mathematically equivalent to Bayes theorem and the likelihood ratio. According to Massaro (1987), the multiplicative integration rule used in the FLMP is an optimal decision rule. In its standard form, the FLMP is applied so as to minimize the differences between obtained and predicted scores, and therefore may be considered more of a *fit* to obtained bimodal scores rather than a *prediction* of optimal bimodal speech performance. This is demonstrated by two consistent aspects of the FLMP. First, the FLMP can only be well fit by iterative manipulations of the unimodal data with full knowledge of the very thing we are trying to predict, namely, the observed auditory-visual recognition results. Briefly, the FLMP seeks to apply the multiplicative integration to unimodal confusion data (i.e., probabilities of responding y given x) to obtain a bimodal prediction, and then compares the predicted result to the obtained AV confusion matrix. To minimize the differences between obtained and predicted scores, an iterative process is used to adjust the obtained unimodal probabilities until a best fit to bimodal performance is obtained. Significantly poorer fits to observed AV data are obtained when only auditory-alone and visual-alone speech recognition data are used as input to the model with no modifications (Massaro, 1998). Second, and perhaps more importantly, human receivers often do better at recognizing consonants than the FLMP predicts. It is therefore difficult to argue that the FLMP predicts optimum performance when human observers can outperform the predicted AV performance, albeit by small amounts.

In contrast, the prelabeling (PRE) model of integration (Braida, 1991) seeks not to optimally fit observed AV data (although the fits are often quite good), but rather seeks to ''label'' incoming bimodal stimuli based on an optimal combination of mutual information gleaned from separate fits to auditory-alone and visual-alone performance. The PRE model first derives an estimate of unimodal information (by iterative multidimensional scaling of the A and V response matrices) and then, using an optimum combination rule, predicts how an unbiased receiver with no interference across modalities might do given the particular unimodal information available. Thus, there appears to be a fundamental difference between the ''optimal'' integration processes proposed by Braida and by Massaro. In our use of the PRE model to date (over 60 hearing-impaired subjects and 15 normal-hearing subjects), predicted AV scores have always equaled or exceeded observed performance. Such was not the case with the FLMP, where subjects' AV scores were equally likely to be underpredicted as to be overpredicted. This result may simply be a by-product of a fitting algorithm that seeks to find the best fit through the data and does not necessarily imply that one model is more ''correct'' than the other. Instead, they simply appear to model different aspects of the integration process. For our purposes (Grant and Seitz, 1998), the PRE model seemed better suited to estimate integration efficiency in an effort to account for individual differences seen in the speech perception of hearing-impaired listeners.

In their recent article, Massaro and Cohen (2000) directly addressed the possibility that under bimodal presentation conditions, some information relative to the unimodal presentation conditions may have been lost, and therefore integration was suboptimal. They modified the original FLMP to allow for a loss of information in the AV condition relative to the auditory and visual conditions. The modification entailed the determination of weighting factors applied when fitting results to the AV condition; one factor describing the possible decrease in auditory support, and one factor describing the possible decrease in visual support. Unimodal weights less than unity represent a loss of information, and therefore, sub-optimal integration. When the modified FLMP was applied to the data from Grant and Seitz (1998), the weights for both A and V conditions were equal to 1 for most subjects and close to 1 for the others, implying no loss of integration efficiency. However, when the modified FLMP was analyzed in terms of five perceptual features (voicing, nasality, frication, place, and duration) rather than the 18 individual consonants, subjects showed a range of weights (i.e., not all weights equal to 1) especially for the visual modality (weights roughly between 0.7 and 1.0). Thus, the two tests applied to the data from Grant and Seitz (1998) resulted in somewhat different outcomes, leaving some confusion concerning the statement that there is ''very little support for the thesis that some individuals might be less-efficient integrators than others'' (Massaro and Cohen, 2000, page 788).

IV. WHAT ARE THE CONSEQUENCES OF BEING A GOOD OR POOR INTEGRATOR?

Assuming that there are substantial differences among subjects in their ability to integrate auditory and visual speech information, what consequences might this have for speech recognition? With the exception of artificially created stimulus situations, such as when auditory and visual speech components are incongruent (e.g., studies of the McGurk effect), it seems reasonable to assume that successful auditory-visual speech integration would imply better overall speech recognition performance than that achievable from auditory-alone or visual-alone input. This is not to suggest that the recognition of all speech tokens would improve with better AV integration, but only that overall speech recognition performance should increase with increases in integration efficiency. The problem with this simple assumption, as mentioned earlier, is that recognition scores can be influenced by a number of factors that have nothing to do with successful or unsuccessful integration. For this reason, Grant and Seitz (1998) have used stepwise multiple regression techniques to determine if there was a significant increase in the amount of variance that could be accounted for when a term representing integration efficiency was added to the list of possible factors. If the amount of recognition variance accounted for increased significantly, then it may be assumed that the proposed measure of integration efficiency was valid. This is precisely what Grant and Seitz found when they based their measure of IE on the optimal processing model proposed by Braida (1991); an additional 10% of the variance in auditory-visual benefit was accounted for (from 43% to 53%) when the IE measure was included along with other factors representing visual and auditory feature extraction and age. Interestingly, there was a tendency for older subjects to be poorer integrators than younger subjects. Given the conclusion by Massaro and Cohen that all subjects integrate with maximal efficiency, one should have expected that the different measures of IE used by Grant and Seitz (susceptibility to McGurk fusion responses, deviations from PRE predictions, and the effects of auditory-visual asynchrony) would have dropped out of the multilinear regression analyses.

V. HOW CAN EXTANT MODELS OF AUDITORY-VISUAL INTEGRATION BE USED TO OBTAIN MEASURES OF IE?

Models of auditory-visual integration appear to be of three basic types. Some models (e.g., Blamey *et al.*, 1989) predict AV speech recognition errors by calculating the joint error probability for the auditory and visual modalities. Thus, given a V error rate of 0.8 (20% correct) and an A error rate of 0.3 (70% correct), the combined AV error rate would be 0.24 (76% correct). Usually subjects make fewer errors than the joint error probability would suggest, possibly setting a lower limit for AV integration predictions. The extent to which subjects exceed this lower limit could then be used as an estimate of IE, with greater excesses indicating better integration. Such an approach was evaluated by Grant *et al.* 1998); however, the correlations between these estimates of IE and overall speech recognition in noise was not significant.

With the FLMP, the assumption is made *a priori* that the model predicts optimal integration (Massaro, 1987; Massaro and Cohen, 2000). As explained earlier, Massaro and Cohen (2000) modified the standard FLMP to include separate weighting factors for auditory and visual input. An estimate of IE is derived from the goodness of fit of the standard FLMP (without auditory and visual weights) compared to the modified FLMP (with auditory and visual weights). This was the approach taken by Massaro and Cohen (2000) in their evaluation of the Grant and Seitz (1998) data sets.

With the PRE model, the assumption is made that the predicted score will optimally reflect the unimodal information with no bias or interference. Thus, all predicted AV scores should be better or equal to obtained AV scores. As mentioned earlier, this has been the case for all the subjects we have tested thus far. The estimate of IE derived from the PRE model, like that described for the model proposed by Blamey *et al.* (1989), is the difference (or alternatively, the ratio) between predicted and obtained AV score. Grant and Seitz (1998) used the deviations between predicted and obtained AV scores as an estimate of IE and showed that there was a significant correlation between IE measures and AV speech recognition in noise and between IE measures and AV benefit (relative and absolute).

VI. WHAT DOES IT MEAN TO HAVE TWO MODELS OF INTEGRATION PRODUCE DIFFERENT RESULTS ON THE SAME DATA SET?

Application of the FLMP and the PRE models to the data set obtained by Grant and Seitz (1998) led to diametrically opposing conclusions. Estimates of IE derived from the PRE model showed that some subjects appeared to be less efficient than others, whereas estimates derived from the FLMP showed that all subjects integrated with equal and near-optimal efficiency. Which conclusion is correct?

The question of which model does a better job at characterizing optimal integration of auditory and visual speech information cannot be answered by referring to the models themselves. Instead, we must first come to an understanding of what we mean when we say that a particular subject is integrating optimally (or not) and what implications, if any, inefficient integration has on speech recognition. Briefly, for stimuli composed of natural auditory-visual speech tokens, being able to integrate auditory and visual cues efficiently implies better overall recognition performance given an accurate accounting of the receiver's auditory and visual capabilities. More precisely, given two individuals with identical auditory and visual speech recognition, the individual with greater IE will do better recognizing auditory-visual speech, and therefore receive more benefit from the addition of visual cues to audition. Unfortunately, it is not possible to match precisely (including the pattern of consonant confusions) different individuals on their unimodal speech recognition performance. Therefore, other tests must be used to determine the relevance of IE measures, such as the multiple regression analyses performed by Grant and Seitz (1998). Although perhaps not a perfect test, the fact that there was an independent means of judging the correctness of the principal conclusion (that the particular choice of IE measure was valid) goes one step further than the modeling approach taken by Massaro and Cohen (2000). In this latter work, no independent test was sought and the value of the modified FLMP as a valid test of IE was simply assumed *a priori*. Even more problematic is the fact that Massaro and Cohen (2000) find some evidence of integration inefficiency when they apply the feature-based model (specifically with regard to a loss of visual information), yet make no attempt to reconcile this with the results of their modality-based model.

VII. CONCLUSIONS

Auditory-visual integration efficiency is a presumed skill employed by subjects independently from their ability to extract information from auditory and visual speech inputs. Although there are no established methods for determining a subject's IE, several proposed solutions to this problem have been reported. One promising methodology employs models of auditory-visual integration to predict optimal AV performance. Differences between model predictions and obtained scores are then used to estimate IE for individual subjects. However, the validity of these derived estimates of integration efficiency cannot be based solely on the accuracy of model fits. Instead, IE must be thought of as a construct supported by other measures, and, it must have consequences with regard to speech recognition performance if there are true differences among individual subjects.

ACKNOWLEDGMENTS

This research was supported by Grant No. DC 00792 from the NIDCD, and by the Clinical Investigation Service, Walter Reed Army Medical Center, under Work Unit No. 2540. I wish to thank Chris Turner and an anonymous reviewer for their helpful comments on an earlier draft of this letter. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

Blamey, P. J., Cowan, R. S. C., Alcantara, J. I., Whitford, L. A., and Clark, G. M. (**1989**). ''Speech perception using combinations of auditory, visual, and tactile information,'' J. Rehabil. Res. Dev. **26**, 15–24.

Braida, L. D. (**1991**). ''Crossmodal integration in the identification of consonant segments,'' Q. J. Exp. Psychol. **43**, 647–677.

Grant, K. W., and Seitz, P. F. (**1998**). ''Measures of auditory-visual integration in nonsense syllables and sentences,'' J. Acoust. Soc. Am. **104**, 2438–2450.

Grant, K. W., and Seitz, P. F. (**2000**). ''The recognition of isolated words and words in sentences: Individual variability in the use of sentence context,'' J. Acoust. Soc. Am. **107**, 1000–1011.

Grant, K. W., Walden, B. E., and Seitz, P. F. (**1998**). ''Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration,'' J. Acoust. Soc. Am. **103**, 2677–2690.

Massaro, D. W. (**1987**). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry* (Erlbaum, Hillsdale, NJ).

Massaro, D. W. (**1998**). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle* (MIT Press, Cambridge, MA).

Massaro, D. W., and Cohen, M. M. (**2000**). ''Tests of auditory-visual integration efficiency within the framework of the fuzzy logical model of perception,'' J. Acoust. Soc. Am. **108**, 784–789.

McGurk, H., and MacDonald, J. (**1976**). ''Hearing lips and seeing voices,'' Nature (London) **264**, 746–748.

Summerfield, Q. (**1987**). ''Some preliminaries to a comprehensive account of audio-visual speech perception,'' in *Hearing by Eye: The Psychology of Lip-Reading*, edited by B. Dodd and R. Campbell (Erlbaum, Hillsdale, NJ), pp. 3–52.

Clicks from Cuvier's beaked whales, *Ziphius cavirostris* (L)

Alexandros Frantzis[a)]
Pelagos Cetacean Research Institute, Terpsichoris 21, 16671 Vouliagmeni, Greece

John C. Goold
Institute of Environmental Science, University of Wales Bangor, Deiniol Road, Bangor, Gwynedd LL57 2UW, United Kingdom

Emmanuel K. Skarsoulis and Michael I. Taroudakis
Foundation for Research and Technology Hellas, Institute of Applied and Computational Mathematics, P.O. Box 1527, GR-71110 Heraklion, Greece

Varvara Kandia
Pelagos Cetacean Research Institute, Terpsichoris 21, 16671 Vouliagmeni, Greece

(Received 13 September 2001; revised 12 March 2002; accepted 13 March 2002)

Recordings of pulsed sounds (clicks) from Cuvier's beaked whales are presented. Such recordings have not been reported in the literature before. Spectrogram analysis of data collected off SW Crete (Greece) from 1998 to 2000 revealed numerous sequences of clicks. Click pulses had durations of about 1 ms and their energy content in the audible spectrum presented a narrow peak between 13 and 17 kHz. Sequences of 35–105 clicks, with duration 15–44 s, were separated by short intersequence pauses of 3–10 s. Interclick intervals appeared fairly constant, primarily oscillating between 0.40 and 0.50 s. Characteristics of Cuvier's beaked whale clicks were consistent with echolocating cetaceans, suggesting that this species do echolocate. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1479149]

PACS numbers: 43.80.Ka [WA]

I. INTRODUCTION

Since 1947, when the bottlenose dolphin was the first cetacean species suggested to detect objects underwater by echolocation, the list of odontocetes shown to echolocate has expanded greatly (Au, 1993). A wide range of species from all the odontocete families have been demonstrated to produce pulsed sounds (commonly called clicks), and it is generally considered that these are used for echolocation purposes (Au, 1993; Richardson *et al.*, 1995).

The family of beaked whales (Ziphiidae) consists of 20 species that account for 28% of all odontocetes (Rice, 1998). However, encounters with most species in the wild are scarce and recordings of their sound emissions are difficult to collect. As a result, their acoustic repertoire is very poorly known. Opportunistic audio recordings from rare sightings or strandings have been reported for six beaked whale species, and among them, only the northern bottlenose whale has been studied systematically (Hooker, 1999).

The only available sound emissions by the Cuvier's beaked whale (*Ziphius cavirostris*), the most common species of the Ziphiid family (Heyning, 1989), is anecdotal (six whistles by Manghi *et al.*, 1999). This lack of data became a serious concern in respect to recent mass strandings of Cuvier's beaked whales that were spatially and temporally correlated with high-level, anthropogenic noise in the local marine environment (Frantzis, 1998; IWC, 2000). Data on both sound production and hearing of beaked whales are of major importance to the understanding of their life habits and interactions with anthropogenic noise. We report the first recordings of Cuvier's beaked whale clicks, and discuss their time and frequency characteristics in the audible range, with respect to their potential echolocation function.

II. METHODS

Acoustic recordings of Cuvier's beaked whales were made from a 16-m vessel off Southwest Crete (Greece, Mediterranean Sea) from 1998 to 2000. Beaked whales were detected visually and were gradually approached. Recordings were started while the whales were at the surface and continued after they dove. In two cases, the proximity of beaked whales was detected acoustically (through the hydrophone) and the recording was started while the whales were diving, before the first visual contact that occurred 15 and 24 minutes later, respectively. During all the recordings the sea state was less than 3 and the sea surface was continuously scanned by naked eye and binoculars, 360 degrees around the research vessel. The time of surfacing and the dive start time of each whale were noted, as well as the visual and/or acoustic presence of other cetacean species. Geographic coordinates were recorded with the aid of a GPS. Bottom depths of sightings were determined by plotting the geographical coordinates on a bathymetric map. Photo-identification of individual whales was accomplished by reviewing miniDV videos recorded at the surface.

The hydrophone array contained two omnidirectional Benthos AQ-4 elements with 30-dB gain preamplifiers. The elements were mounted 3 m apart along the axis of a 10-m oil-filled polyurethane tube. The frequency response of the elements was flat ±1.5 dB and flat ±2.0 dB for the 1 Hz to

a)Electronic mail: afrantzis@otenet.gr

TABLE I. Data regarding the tape segments of Cuvier's beaked whale recordings selected for interclick interval (ICI) analysis.

Date Time	Duration (No. of clicks)	No. of whales	Bottom depth (m)	Context
2 July 2000 13:28:58	71.7 s (119)	2	1020	Middle of a long dive of 68 min started at 12:55:30
31 July 2000 12:21:47	101.8 s (177)	2	500	Beginning of a long dive started at 12:19:16
1 August 2000 11:57:53	107.3 s (227)	1	1300	During a dive started at 11:49:50

15 kHz and 15–25 kHz bandwidths, respectively. The array was towed 100 m behind the vessel or, when the vessel was not in motion, was left to sink into a vertical position 100 m below the stern. Signals from the array were high-pass filtered (200 Hz filter break, −12 dB per octave roll off) before being recorded with a Sony TCD-D8 digital tape recorder (DAT), at 48 kHz sampling frequency. This DAT recorder has a flat frequency response ±1.0 dB for the bandwidth 20 Hz to 22 kHz, and cuts off above 24 kHz. As a result, possible higher frequency components of the Cuvier's beaked whale clicks could not be recorded. Similarly, it was not possible to obtain the source level and position of the sound-producing animals relative to the hydrophone.

Preliminary waveforms and spectrograms were made using BatSound 1.2 and Sound Forge XP 4.0. Only very short parts of the recordings contained clearly audible Cuvier's beaked whale clicks, although visually clear clicks could be identified in the waveforms and spectrograms for sequences as long as 14 min. In a total of 5 h and 3 min of recordings made in proximity of Cuvier's beaked whales, sound emissions of this species were detected in 59 min. In order to avoid contamination of our data with sounds from other species, we processed only tape segments for which no dolphins or other cetaceans had been sighted or heard for one hour before or after the recordings. Nine tape segments were selected for final analysis because only one whale was recorded and had high signal-to-noise ratio. These segments started 2.5 to 34.0 min after the start of long dives. Three segments contained clicks that could be identified and tracked with no interruption for periods longer than 70 s, and were selected for the interclick interval (ICI) analysis (Table I). These click trains originated from three different individuals (on the basis of photo-identification results), encountered on three different days.

Clicks appeared as narrow spectrogram peaks at around 15 kHz, in most cases followed by a surface echo coming a few tens of milliseconds later. Click intensity was variable over a period of tens of seconds, and only short sections showed high signal level. These sections were used for the frequency and pulse duration analysis. The sound files were imported into MATLAB for detailed inspection of the individual pulse shapes and spectra. Only one channel (the right) was used for the analysis. A digital high-pass filter set at 500 Hz was applied to reduce noise. The onset of each individual click was marked to the nearest sample point, with an accuracy of about 0.02 ms. Similarly, a marker was laid down at the position where the pulse decayed into the background noise. These marker data were subsequently compiled to yield ICIs, defined as the time difference between the onset markers of two successive clicks, and pulse durations, defined as the time difference between onset and decay markers of each click. Onset markers also served as reference points for spectral analysis targeted at the short sections of data containing the click pulses. Each click was centered in a 256-point FFT window (duration 5.3 ms) and the section was modulated with a Hanning window to remove edge effects.

III. RESULTS

Pulse duration of Cuvier's beaked whale clicks ranged from 0.7 to 1.6 ms, with an average of 1.08 ms (s.d. $=0.26$, $n=142$). The energy of the clicks was concentrated into a narrow peak between 13 and 17 kHz (Fig. 1). Spectral analysis of sequential clicks revealed, repeatedly, a progressive slide of this narrow-band peak, between 13 and 16 kHz, during the course of some 20 clicks, in less than 10 s. However, distortions due to possible off-axis recording cannot be excluded, since the position of the whales relative to the hydrophones was not known.

The analysis of the three segments that contained long series of clicks (Table I) revealed that click production in Cuvier's beaked whales is not continuous. Sequences of 35–105 clicks, with duration 15.5–44.5 s, were separated by short intersequence pauses (periods of silence) of 3.0–10.3 s. This pattern was obvious in all analyzed segments, which contained seven complete, and three incomplete click sequences in total, with eight pauses among them [Figs. 2(a)–(c)]. As complete click sequences we define those that were recorded from their first to their last click and consequently were preceded and followed by a pause. All pauses were

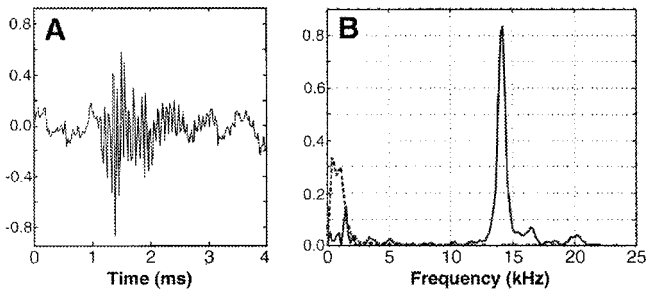


FIG. 1. Typical waveform in background noise (a) and spectral density (b) in the audible frequency range of Cuvier's beaked whale clicks. The dashed line in the spectrogram represents the level of background noise (high-pass filtered at 500 Hz).

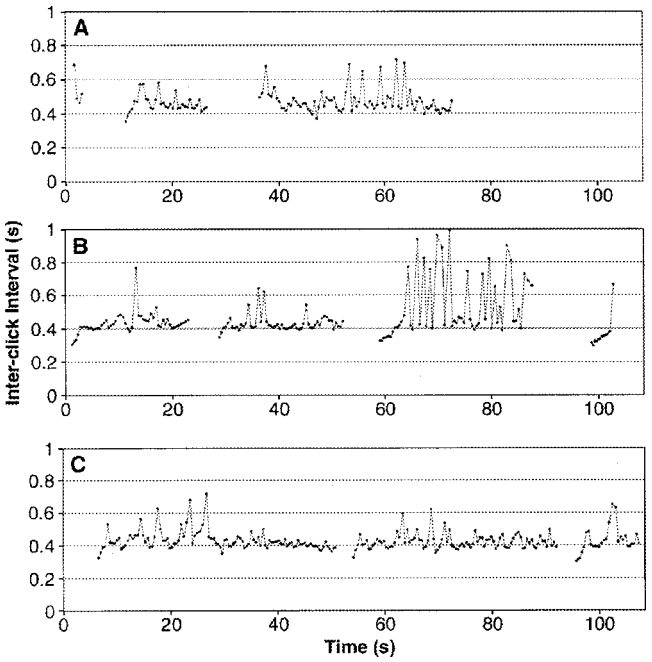


FIG. 2. Variation in interclick interval (ICI) during the dive of three different Cuvier's beaked whales [plots (a)–(c)], recorded on three different days. Plots contain: one incomplete and two complete sequences (a), three complete and one incomplete sequence (b), a single click at time 0 s followed by a pause, two complete, and one incomplete sequence (c).

three to ten times longer than the maximum ICI encountered, and were discarded from any ICI statistical analysis.

The ICI appeared fairly constant, oscillating between 0.40 and 0.50 s for most parts of seven of the eight complete click sequences [Figs. 2(a)–(c)]. Some sharp oscillations were also present. In one click sequence [third in Fig. 2(b)], although the ICI baseline was at about 0.40 s, a sharp oscillation was the dominant pattern, with the highest ICIs ranging between 0.70 and 1.00 s. The changes from the baseline to the next maximum and back to the baseline were not gradual, but occurred from one ICI to the next. In total, the ICI ranged from 0.295 to 0.989 s. Although the differences between average ICIs from the three analyzed segments were significant (single factor ANOVA, $n=512$, $F=11.29$, $p<0.001$), the three ICI distributions had similar modes (0.40–0.45 ms), and more than 75% of ICIs ranged between 0.35 and 0.50 ms in all three cases. The overall average ICI was 0.444 s ($n=512$, s.d.$=0.092$).

All click sequences presented a common starting pattern: very short ICIs increasing progressively [Figs. 2(a)–(c)]. In all but one case, the first ICI of each sequence had the minimum value of the entire click sequence. No obvious trend was observed for the last clicks of click sequences.

IV. DISCUSSION

Efforts to record Cuvier's beaked whale clicks have been made in the past (Dawson *et al.*, 1998; Manghi *et al.*, 1999). However, their lack of success raised the possibility that Cuvier's beaked whales were substantially less ''vocal'' than other beaked whales (Dawson *et al.*, 1998). Our repeated recordings of Cuvier's beaked whale sounds indicate that this species produces clicks as often as other odontocetes. These clicks had not been detected up to now, for reasons likely common to most beaked whale species.

While finding free-ranging Cuvier's beaked whales is difficult because of their behavior, we believe that the main reason why no recordings have been previously reported is that Cuvier's beaked whale clicks are rarely audible to most humans when heard or recorded through conventional omnidirectional hydrophones. Both their frequency and short pulse duration make them difficult to detect by ear. In the audible frequency range, their only significant components above background noise were between 13 and 17 kHz, just at, or above the limit of useful sensitivity for most humans. In addition, the sound level of clicks was low in most parts of our recordings, with loud clicks occurring intermittently for only a few seconds. Cuvier's beaked whales may also produce ultrasounds, as is the case with northern bottlenose whales (Hooker, 1999) and Baird's beaked whales (Dawson *et al.*, 1998). However, in this work only frequency components in the audible spectrum could be recorded and analyzed with the available instrumentation.

The oscillating signal levels, as well as the progressive slide of the recorded peak frequency, suggest that Cuvier's beaked whale clicks are directional, and that the hydrophone may be picking up different parts of the beam as the whale changes orientation underwater. If this is the case, the energy peaks of clicks recorded off-axis are biased towards lower frequencies, since the latter give rise to a broader beam pattern. We have observed and recorded on video one Cuvier's beaked whale just under the surface, changing the direction of its head (right–left–right, etc.) nine times, as if ''scanning'' the vessel, while coming towards it. The mean scanning angle was about 50° and the mean rate of four complete scans was one scan per 2.2 s. Unfortunately, no audio recording was made during this ''scanning'' behavior.

The average interclick interval of diving Cuvier's beaked whales (0.44 s) is almost identical to that reported by Hooker (1999) for ''distant clicks'' of diving northern bottlenose whales (0.4 s). ICIs of regular clicks from foraging sperm whale oscillate around 1 s (Gordon, 1987; Goold and Jones, 1995; Douglas, 2000). For both sperm whales and northern bottlenose whales there is evidence that they produce clicks for echolocation purposes (Gordon, 1987; Hooker, 1999; Møhl *et al.*, 2000; Jaquet *et al.*, 2001). Both species are mainly teuthivorous deep divers (Rice, 1989; Mead, 1989; Hooker and Baird, 1999) with diets very similar to Cuvier's beaked whales (Heyning, 1989).

On the basis of the above similarities it appears that the clicks recorded from Cuvier's beaked whales are also used for echolocation as has been assumed or demonstrated for most odontocetes (Au, 1993; Richardson *et al.*, 1995). If this is the case, the ICI should approximate the two-way transit time to the target that the whale is echolocating on, or the maximum detection range at which searching is taking place (Au, 1993). An ICI of 0.44 s implies a searching range of about 310 m ($c_{\text{water}}=1500\ \text{m s}^{-1}$). This is practically equal to what has been estimated for northern bottlenose whales (Hooker, 1999), but about half that estimated for sperm whales (Goold and Jones, 1995). Furthermore, the increasing ICIs in the beginning of all analyzed click sequences indicate

that echo-searching in Cuvier's beaked whales commences from about 210 m to expand to about 310 m in the next 1–3 s. A similar pattern was recently observed in click sequences from deep diving sperm whales; 60% of click sequences begin with increasing ICIs and only 3% with decreasing ICIs (Frantzis *et al.*, in preparation).

Small and moderate ICI oscillations, such as those in most click sequences of Fig. 2, have already been reported for echolocating bottlenose dolphins, and were characterized as ''typical'' of all click trains examined in the presence as well as in the absence of a target (Au, 1993). Similar oscillations have been observed in the ICIs of diving sperm whales and various explanations have been proposed (Goold and Jones, 1995; Douglas, 2000). The sharp oscillations observed in the last complete click sequence of Fig. 2(b) seem unusual and their potential function is unclear.

The detected intersequence pauses (short periods of silence) in Cuvier's beaked whale recordings present an astonishing similarity with sperm whale acoustic behavior during deep dives (Gordon 1987; Douglas, 2000), although at a slightly different time scale. Sperm whale pauses last 2.5 to 58 s, and occur after creaks, or after regular click sequences of 0.5 to 4.4 min (Frantzis *et al.*, in preparation). Both pauses and click sequences from Cuvier's beaked whales (respective durations 3.0–10.3 s and 15.5–44.5 s) were many times shorter than those of sperm whales. The number of clicks between two pauses varies twice as much in sperm whales as in Cuvier's beaked whales (ranges 18–215 and 35–105, respectively). The purpose of regular intersequence pauses between click trains is not yet understood. They may represent the result of feeding after the capture of a prey, a period with no need for sound production, or a necessary, short resting period for the click-producing organ(s). In any case, their detection in Cuvier's beaked whales shows that they are not a sperm whale peculiarity. We detected no sounds similar to sperm whale creaks (rapid bursts of clicks with repetition rates of up to 200 s^{-1}) in Cuvier's beaked whales. Constant ICIs and absence of creaks were also reported for northern bottlenose whales and led to the hypothesis that this species could use vision in the final stages of a fairly passive prey capture (Hooker, 1999; Hooker and Baird, 1999). This hypothesis may also apply to Cuvier's beaked whales, since all three prey species found in the stomachs of eight specimens stranded in Greece have photophores (Lefkaditou and Poulopoulos, 1998; Frantzis *et al.*, in preparation). However, more data are needed before concluding that creaks are absent in this species.

Given our limited data, we can do little more than speculate about the precise function of the recorded clicks. However, the fact that diving Cuvier's beaked whales emit clicks with characteristics that are consistent with echolocating cetaceans suggests that Cuvier's beaked whales also echolocate.

AKNOWLEDGMENTS

The first author wants to thank the International Fund for Animal Welfare (IFAW) for a grant in support of the data analysis and preparation of this paper and for their general support to novel and conservation-focused scientific research regarding whales. Special thanks to A. Moscrop and T. Lewis who supported the idea and necessity of this grant. J. C. Goold would like to thank B. Rossiter and Cetacean Society International for a travel grant in support of a working visit to the Pelagos Cetacean Research Institute. We also wish to thank C. Menhenett for providing hydrophones at cost price, R. Swift and D. Gillespie for helping to set up the Cretan Project; P. Alexiadou for her participation in the fieldwork and in the organization of the cetacean acoustic database of Pelagos Institute; and all members and ecovolunteers of the Pelagos Institute who participated in the Cretan Project and supported its expeditions.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).
Dawson, S., Barlow, J., and Ljundberg, D. (**1998**). ''Sounds recorded from Baird's beaked whale, *Berardius bairdii*,'' Mar. Mammal Sci. **14**, 335–344.
Douglas, L. (**2000**). ''Click counting: An acoustic censuing method for estimating sperm whale abundance,'' M.Sc. thesis, University of Otago, Dunedin, New Zealand.
Frantzis, A. (**1998**). ''Does acoustic testing strand whales?'' Nature (London) **392**, 29.
Frantzis, A. *et al.* (in preparation).
Goold, J. C., and Jones, S. E. (**1995**). ''Time and frequency domain characteristics of sperm whale clicks,'' J. Acoust. Soc. Am. **98**, 1279–1291.
Gordon, J. C. D. (**1987**). ''The behavior and Ecology of Sperm Whales off Sri Lanka,'' Ph.D. thesis, University of Cambridge.
Heyning, J. E. (**1989**). ''Cuvier's Beaked Whale *Ziphius cavirostris* G. Cuvier, 1823,'' in *Handbook of Marine Mammals Volume 4: River Dolphins and the Larger Toothed Whales*, edited by S. H. Ridgway and R. Harrison (Academic, London), pp. 289–308.
Hooker, S. K. (**1999**). ''Resource and habitat use of northern bottlenose whales in the Gully: ecology, diving and ranging behavior,'' Ph.D. thesis, Dalhousie University.
Hooker, S. K., and Baird, R. W. (**1999**). ''Deep-diving behavior of the northern bottlenose whale, *Hyperoodon ampullatus* (Cetacea: Ziphiidae),'' Proc. R. Soc. London, Ser. B **266**, 671–676.
International Whaling Committee (**2000**). ''Report of the Standing Committee of Environmental Concerns,'' in Report of the Scientific Committee of the International Whaling Commission SC/52, p. 28.
Jaquet, N., Dawson, S., and Douglas, L. (**2001**). ''Vocal behavior of male sperm whales: Why do they click?'' J. Acoust. Soc. Am. **109**, 2254–2259.
Lefkaditou, E., and Poulopoulos, Y. (**1998**). ''Cephalopod remains in the stomach-content of beaked whales, *Ziphius cavirostris* (Cuvier, 1823), from the Ionian Sea,'' Rapp. Comm. Int. Mer Médit. **35**, 460–461.
Manghi, M., Montesi, G., Fossati, C., Pavan, G., Priano, M., and Teloni, V. (**1999**). ''Cuvier's beaked whales in the Ionian Sea: First recordings of their sounds,'' Eur. Res. Cetaceans **13**, 39–42.
Mead, J. G. (**1989**). ''Bottlenose Whales *Hyperoodon ampullatus* (Forster, 1770) and *Hyperoodon planifrons* Flower, 1882,'' in *Handbook of Marine Mammals Volume 4: River Dolphins and the Larger Toothed Whales*, edited by S. H. Ridgway and R. Harrison (Academic, London), pp. 321–348.
Møhl, B., Wahlberg, M., Madsen, P. T., Miller, L. A., and Surlykke, A. (**2000**). ''Sperm whale clicks: Directionality and source level revisited,'' J. Acoust. Soc. Am. **107**, 638–648.
Rice D. W. (**1989**). ''Sperm Whale *Physeter macrocephalus* Linnaeus, 1758,'' in *Handbook of Marine Mammals Volume 4: River Dolphins and the Larger Toothed Whales*, edited by S. H. Ridgway and R. Harrison (Academic, London), pp. 177–233.
Rice D. W. (**1998**). *Marine Mammals of the World. Systematics and Distribution*, Special publications of the Society for Marine Mammalogy, Vol. 4 (Allen, Lawrence), pp. 1–231.
Richardson, W. J., Greene, C. R. Jr., Malme C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).

Floquet wave homogenization of periodic anisotropic media

L. Wang and S. I. Rokhlin[a)]
The Ohio State University, Nondestructive Evaluation Program, Edison Joining Technology Center, Columbus, Ohio 43221

(Received 12 June 2001; revised 14 February 2002; accepted 30 April 2002)

A dynamic homogenization method based on Floquet wave theory is developed. The theory is based on the equivalency within the homogenization domain of Floquet waves in a periodic anisotropic medium and plane waves in a dispersive homogeneous anisotropic medium. A simple procedure has been developed to estimate analytically critical angles and the upper frequency bound of this homogenization domain. Using this method, the frequency dependent effective elastic constants are obtained and examples for [0/90] and [0/45/90/−45] composites are given. By comparison with an exact theory, it is shown that the time domain signal propagation in a periodic laminate is well described by the Floquet wave homogenization theory in the homogenization domain. It is also shown that in the static limit the results are identical to those calculated by static homogenization theory (the generalized method of cells). The potential applications of the method are discussed. [DOI: 10.1121/1.1488942]

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

I. INTRODUCTION

One of the important applications of elastic wave propagation in layered anisotropic media is the ultrasonic characterization of advanced multilayered composites. The widely used multiply graphite/epoxy composites are stacks of orthotropic laminas (plies) made of 5 μm diameter graphite fibers unidirectionally embedded in epoxy resin. These laminas are extremely anisotropic with max/min longitudinal wave velocity ratio about three. The laminas are assembled in different periodic stacking sequences of fiber direction (for example 0°/45°/90°/−45°) to form multilayered, and often very thick, structures.

Due to the importance of this material for structural applications, significant research has been devoted to elastic wave propagation in such materials. Extensive studies of wave propagation in laminated composites was started in the 1960s by Achenbach and co-workers[1–3] considering the laminate layers as isotropic. They described spectra of guided and Floquet waves and discuss a low frequency homogenization using static effective moduli theory. This work has been reviewed by Achenbach.[4] Datta and co-workers[5–9] performed comprehensive studies of wave propagation in multiple composites using a thin layer discretization method. The low-frequency-dependent effective elastic properties of a periodic anisotropic medium is considered in Ref. 6. Dispersion of elastic waves in crossply composites was investigated in detail in Ref. 7; comparison of the static effective medium approximation to exact calculation showed that the approximation is good when the number of plies is quite large. Otherwise, it is good only for the lowest modes at low frequencies. It was also shown[8] for transient wave propagation in a periodic [0/90] crossply plate that the effective medium approximation compares better with exact calculation at low frequencies and when the number of plies is large. A comprehensive review of the dispersion of different modes in composite plates and comparison with the effective static elastic property approximation can be found in Ref. 9. Several studies on wave propagation in composites are reported in Ref. 10. Nayfeh[11,12] developed a transfer matrix method for general multilayered anisotropic media, obtained the Floquet wave equation for the periodic case and considered several numerical examples. Experimental and theoretical examples for multidirectional composites are given by Chimenti and Nayfeh[13] and Nayfeh and Chimenti.[14] Stable reformulation of the transfer matrix method with applications to multidirectional composites is given in Ref. 15. Braga and Herrmann[16] also addressed periodic arbitrarily anisotropic media with a transfer matrix method using the Stroh formalism. They performed a very detailed analysis of the Floquet wave characteristic equation and studied stop and pass bands (Brillouin zones). Potel and Belleval[17] applied the Floquet wave solution to finite periodic media and specifically addressed composite media homogenization using Floquet wave theory. They obtained results equivalent to low frequency homogenization used in mechanics of composite materials. Schoenberg[18] discussed homogenization of layered media of alternating isotropic fluid and solid layers. Effective-media results for multilayered composites have also been obtained by Hosten[19] using stiffness matrix invariants and static effective theory. Shull *et al.*[20] showed that the special Lamb wave dispersion behavior is caused by the pass and stop bands of the Floquet waves in a plate containing four layers of aluminum and three layers of aramid-epoxy composite lamina. Using the Floquet wave concept, Auld *et al.*[21] showed that there exists a new class of horizontal shear surface waves on a laminated composite. Rokhlin *et al.*[22] and Wang and Rokhlin[23] have demonstrated the relation between the transmission amplitude distribution and the Floquet wave pass and stop bands for multilayered composites.

The main objective of this paper is to develop a dynamic

a)Author to whom correspondence should be addressed. Electronic mail: rokhlin.2@osu.edu

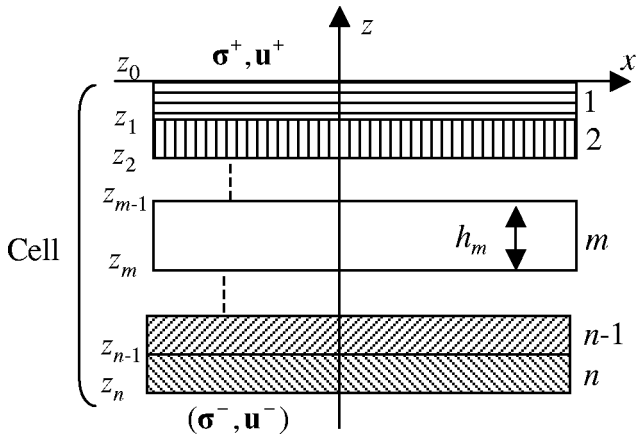


FIG. 1. A multilayered composite and coordinate system.

Floquet wave homogenization method for a periodic layered anisotropic medium with principal application to composites. It is done by equating at a given frequency the slowness surface of Floquet waves propagating in an actual medium to the slowness surface of the bulk waves propagating in the effective dispersive anisotropic homogeneous medium. The frequency-dependent effective elastic constants are obtained from optimization of the Christoffel equation solution by least-squares fit of the Floquet wave slowness surface. The dynamic homogenization method is described in Sec. II. We derive the Floquet wave characteristic equation using the recursive stiffness matrix method for periodic media with arbitrary anisotropy, which is computationally stable and efficient.[15] In Sec. III we provide numerical illustrations of the Floquet wave homogenization method and discuss its accuracy. In Sec. IV we give a simple approximate analysis of the upper frequency bound for the homogenization domain and provide a simple criterion for its determination.

II. FLOQUET WAVE-BASED DYNAMIC HOMOGENIZATION: ALGORITHM

A. Floquet wave characteristic equation and velocity

The Floquet wave spectrum of multilayered composites (stop and pass bands) and slowness surfaces have been extensively investigated in Refs. 16 and 17. To obtain the Floquet wave characteristic equation one should apply the periodicity condition.[12,16]

$$\begin{bmatrix} \mathbf{u}^+ \\ \boldsymbol{\sigma}^+ \end{bmatrix} = \mathbf{I} e^{i\zeta h} \begin{bmatrix} \mathbf{u}^- \\ \boldsymbol{\sigma}^- \end{bmatrix}, \tag{1}$$

where ζ represents the Floquet wave number and h the thickness of the unit cell. Different methods can be used to obtain the characteristic equation for ζ; among them the transfer matrix method is often used.[12,16,17] In our work, for Floquet wave computation we utilized a stiffness matrix method. Let us consider an anisotropic cell $[\theta_1/\theta_2/\cdots\theta_{n-1}/\theta_n]$, consisting of n arbitrarily oriented anisotropic layers as illustrated in Fig. 1. The cell stiffness matrix $\mathbf{K}_c$,

$$\begin{bmatrix} \boldsymbol{\sigma}^+ \\ \boldsymbol{\sigma}^- \end{bmatrix} = \begin{bmatrix} \mathbf{K}_c^{11} & \mathbf{K}_c^{12} \\ \mathbf{K}_c^{21} & \mathbf{K}_c^{22} \end{bmatrix} \begin{bmatrix} \mathbf{u}^+ \\ \mathbf{u}^- \end{bmatrix} \tag{2}$$

which relates the stresses and displacements at cell top $(\boldsymbol{\sigma}^+, \mathbf{u}^+)$ and bottom $(\boldsymbol{\sigma}^-, \mathbf{u}^-)$ surfaces, is obtained from the layer stiffness matrices $\mathbf{K}_m$ (A4) by the recursive algorithm (A5) (Appendix A).

Using the cell stiffness matrix [Eq. (2)] and periodicity condition [Eq. (1)], we obtain the system of equations for the Floquet wave displacements

$$(e^{i\zeta h}\mathbf{K}_c^{21} - e^{-i\zeta h}\mathbf{K}_c^{12} + \mathbf{K}_c^{22} - \mathbf{K}_c^{11})\mathbf{u}^- = 0. \tag{3}$$

$\mathbf{u}^-$ is the Floquet wave unit displacement vector and is equivalent to the displacement polarization $\mathbf{p}$. Equation (3) is equivalent to the Christoffel equation for plane waves in a homogeneous anisotropic medium. The characteristic equation for the Floquet waves in an arbitrarily anisotropic medium is obtained by equating the determinant of the system (3) to zero. If the z axis is a material symmetry axis, then this characteristic equation can be simplified to

$$A_3\cos(3\zeta h) + A_2\cos(2\zeta h) + A_1\cos(\zeta h) + A_0 = 0, \tag{4}$$

where

$$A_3 = |\mathbf{K}_c^{21}|,$$

$$A_2 = (|\mathbf{M} + \mathbf{K}_c^{21}| + |\mathbf{M} - \mathbf{K}_c^{21}|)/2 - |\mathbf{M}|,$$

$$A_1 = (|\mathbf{K}_c^{12} + \mathbf{K}_c^{21}| + |\mathbf{K}_c^{12} - \mathbf{K}_c^{21}| - |\mathbf{K}_c^{12} + \mathbf{M}| - |\mathbf{K}_c^{12} - \mathbf{M}|)/2 + |\mathbf{K}_c^{12}| - |\mathbf{K}_c^{21}|,$$

$$A_0 = 2|\mathbf{M}| + (|\mathbf{M} + \mathbf{K}_c^{12} - \mathbf{K}_c^{21}| + |\mathbf{M} + \mathbf{K}_c^{12} - \mathbf{K}_c^{21}| - |\mathbf{K}_c^{12} + \mathbf{M}| - |\mathbf{K}_c^{12} - \mathbf{M}| + |\mathbf{K}_c^{21} + \mathbf{M}| - |\mathbf{K}_c^{21} - \mathbf{M}|)/2,$$

$\mathbf{M} = \mathbf{K}_c^{22} - \mathbf{K}_c^{11}$; $|\mathbf{M}|$ represents the determinant of matrix $\mathbf{M}$. Equation (4) is similar to that obtained by Nayfeh,[12] but the coefficients are represented through the stiffness matrix elements rather than the transfer matrix. Equation (4) has six independent solutions for ζ $(-\pi < \zeta h < +\pi)$. These solutions satisfy $\zeta_1 = -\zeta_2$, $\zeta_4 = -\zeta_3$, $\zeta_5 = -\zeta_6$.

The pass and stop band phenomena and slowness surfaces for graphite/epoxy crossply composites have been investigated by Braga and Herrmann[16] and Potel *et al.*[17] To facilitate further discussion we consider examples of the Floquet wave spectrum for [0/90] and [0/45/−45/90] composites with the incident plane along the 0° direction in the top ply. The spectra are shown in Fig. 2 in the $k_x - \omega$ domains and the lamina properties used for computation are given in Table I. To relate this to ultrasonic experiments, we convert k_x to an equivalent incident angle from water using $k_x = \omega \sin(\theta)/V_0$ (V_0 is the acoustic velocity in water 1.5 mm/μs). The gray level represents the number of propagating Floquet waves in the pass band. White indicates the area where all three Floquet waves have real wave numbers and black corresponds to the nonpropagating region (stop bands) for all three Floquet waves (they become nonhomogenous with complex wave numbers). Two different levels of gray between white and black indicate that one or two of the Floquet waves have real wave numbers and are propagating waves. As shown in Fig. 2, at low frequencies below the vertical dashed line at about 2.2 MHz, the stop and pass

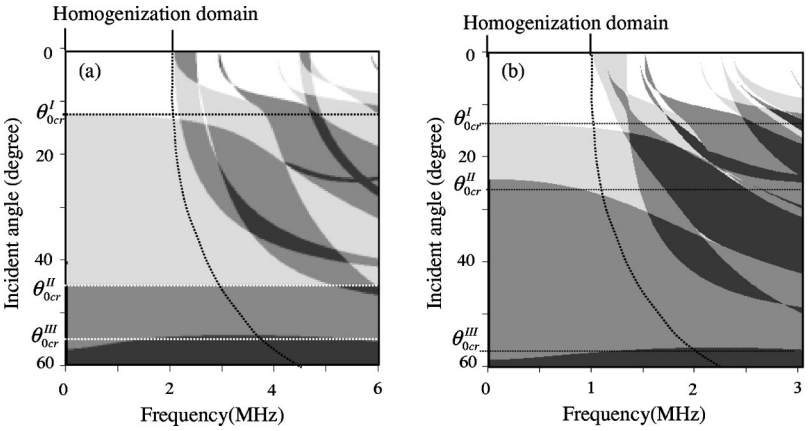


FIG. 2. Pass and stop bands of the three Floquet waves in k_x and ω domains for (a) [0/90] composites and (b) [0/45/90/−45] composites. The propagating plane is oriented along 0° (white, three propagating waves; gray, one or two propagating waves; black, stop band no propagating waves permitted).

bands separate into four regimes divided by three critical incident angles. With increase of incident angle, propagating Floquet waves become nonhomogeneous after reaching the Floquet wave critical angle. This behavior is similar to waves propagating in an anisotropic homogeneous medium with three critical angles. The Floquet wave critical angles are slightly dependent on frequency. The smallest critical angle decreases as the frequency increases.

In this domain (below 2.2 MHz), we may consider the periodic medium as an effective homogeneous anisotropic medium supporting three plane waves, which become nonhomogeneous waves above the critical angles. Estimation of the boundaries of this homogenization domain and determination of the effective medium elastic properties are the objectives of this paper and will be discussed in the following sections.

B. Homogenization approach

As discussed in the preceding section, at frequencies below the first stop band the Floquet waves behave as plane waves in homogeneous anisotropic media. This forms the basis for our homogenization approach. To replace the actual multilayered cell, we define the effective medium cell as an anisotropic homogeneous cell for which the stiffness $\mathbf{K}_{\text{eff}}$ and transfer $\mathbf{B}_{\text{eff}}$ matrices are equal to the stiffness and transfer matrices of the actual cell

$$\mathbf{K}_c = \mathbf{K}_{\text{eff}}; \quad \mathbf{B}_c = \mathbf{B}_{\text{eff}}. \tag{5}$$

Since all effective cells are homogeneous and identical, after such replacement the whole periodic anisotropic medium transforms to an anisotropic homogeneous medium. This replacement (substitution of the actual cell for a homogeneous effective cell) does not change the Floquet wave velocities ($V_F = V_{F_{\text{eff}}}$) or other properties since, due to the equality of the stiffness matrices, the stresses and displacements on the cell boundaries do not change and the boundary conditions (1) and Eq. (4) are satisfied exactly. It is shown in Appendix B that for an arbitrary frequency and propagation direction in a homogeneous anisotropic medium the Floquet wave wave numbers (velocities) are equal to the wave numbers of plane waves (velocities) in this medium calculated from the Christoffel equation. However, for a periodic medium, whose cell is constituted by n laminas, the homogenization requirement (5) and thus the equality of Floquet wave numbers in the periodic medium to the wave numbers in the effective homogeneous anisotropic medium is satisfied locally; i.e., in a general case the condition (5) may be satisfied only for a specific frequency and a specific direction. Therefore for the actual anisotropic periodic medium one obtains different effective media for different propagation directions and frequencies. To produce realistic effective anisotropic media we need to require a single Christoffel tensor for the effective medium at a given frequency. For this we equate at a given frequency the slowness surface for the Floquet waves in the actual medium with the slowness surface for the bulk wave in an effective dispersive anisotropic medium. The frequency effect will be accounted for by dispersion, i.e., frequency dependency of elastic constants. The frequency domain where this can be done we call the homogenization domain.

TABLE I. Lamina properties.

Elastic constants of one lamina (GPa)	
C_{11}	143.2
C_{22}	15.8
C_{12}	7.5
C_{23}	8.2
C_{55}	7.0
Density (g/cm^3)	1.6
Thickness (mm)	0.194

C. Dynamic elastic constants of effective homogenous anisotropic media

The elastic constants of a homogeneous anisotropic medium can be determined[24–26] from velocity measurements in different propagation directions. To demonstrate the application of the homogenization method, we use the Floquet wave theory to calculate the velocities in different propagation directions and equate them to the plane wave velocities of the effective homogeneous medium, then using these velocities and the same procedure as in Ref. 24, we ''reconstruct'' the elastic constants for the effective homogeneous medium.

Figure 3 outlines the algorithm to obtain the effective elastic constants of a homogenized periodic medium from the Floquet wave slowness surface. The Floquet wave velocities $V_i^F(\theta_i)$ are calculated from the Floquet wave characteristic equation [Eq. (4)] in different propagation directions

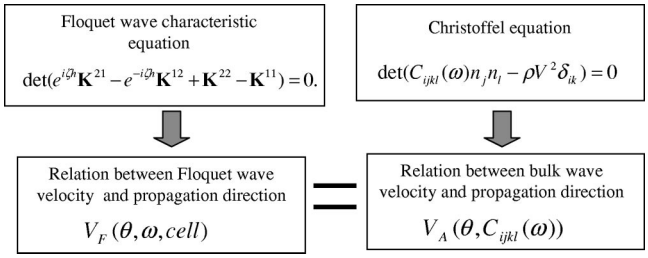


FIG. 3. Homogenization procedure.

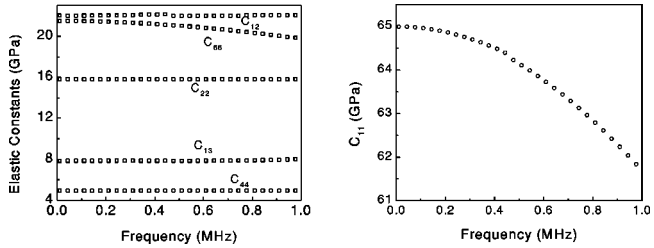


FIG. 5. Effective elastic constant C_{ij} for [0/45/90/−45].

θ_i. The plane wave velocities $V_i^{\text{eff}}(\theta_i)$ for an effective homogeneous medium are determined from the Christoffel equation [Eq. (A2)] and are equated to the Floquet wave velocities in the same direction. The set of effective elastic constants is obtained from the plane wave velocities using inversion of the Christoffel equation. To do this, we select an arbitrary set of elastic constants of the effective medium as initial guesses and calculate the velocities for the homogeneous medium using the Christoffel equation (A2). Since the initial guesses for the effective elastic constants are arbitrary, the velocities $V_i^{\text{eff}}(\theta_i)$ in the effective homogeneous medium are different from the Floquet wave velocities $V_i^F(\theta_i)$ in the periodic medium. The unknown effective elastic constants are obtained by a least-squares nonlinear minimization[24,25]

$$\min_{C_{ijkl} \in R^n} \frac{1}{2} \sum_{i=1}^{m} (V_i^F(\theta_i) - V_i^{\text{eff}}(\theta_i))^2, \tag{6}$$

where $V_i^F(\theta_i)$ is the set of calculated Floquet wave velocities propagating at angles θ_i, $V_i^{\text{eff}}(\theta_i)$ are the calculated wave velocities for an effective medium, n is the number of independent parameters to be determined which depends on the cell structure, m is the number of velocity values for different propagation directions θ_i used. This algorithm converges to the correct velocities independently of the initial guesses.[25]

To determine the effective elastic constants, one has to know the symmetry of the effective media and the number of independent elastic constants. The symmetry is obtained from the symmetry of the Floquet wave slowness curves. For a [0/90] crossply composite, the cell has [0/90] symmetry and therefore the 0°, 90°, and 45° orientations are the symmetry planes. This can be easily seen from the stop and pass bands distribution in k_x and the rotation angle domain calculated using Eq. (4). This symmetry can be matched by a homogeneous medium with higher tetragonal symmetry (with $C_{16}=C_{26}=0$) which has six independent elastic constants and can be considered as a subset of an orthotropic material. For a quasi-isotropic composite [0/45/90/−45], in the homogenization frequency domain, the Floquet wave slowness surface in k_x and the rotation angle domain is independent of rotation angle and therefore the 1–2 plane (the plane of cell boundaries) is a plane of isotropy; thus the material can be modeled as a transversely isotropic medium with five independent elastic constants.

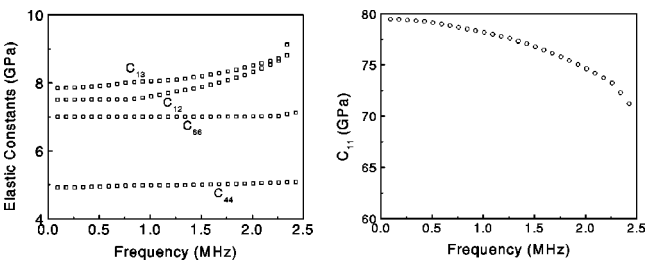


FIG. 4. Effective elastic constant C_{ij} for [0/90].

III. COMPUTATIONAL EXAMPLES

A. Determination of effective elastic constants and comparison with stiffness for static effective media

As an example, we will use [0/90] and [0/45/90/−45] composites to calculate the effective elastic constants using the algorithm described above. Using Eq. (4), the Floquet wave slowness surfaces are generated at 10 different wave propagation planes rotated by angle φ around the symmetry axis z (3) (Fig. 1). In each propagation plane, 40 different propagation angles θ are used. Equation (A2) is used to calculate the plane wave slowness surface in the effective homogeneous medium in the same propagation directions. Therefore, a total of $m=10\times 40=400$ points are used at one frequency in the minimization equation (6).

The optimization algorithm is applied to obtain the frequency-dependent effective elastic constants. Figures 4 and 5 display the effective elastic constants as function of frequency for composites [0/90] and [0/45/90/−45], respectively. As shown in Figs. 4 and 5, C_{11} decreases as frequency

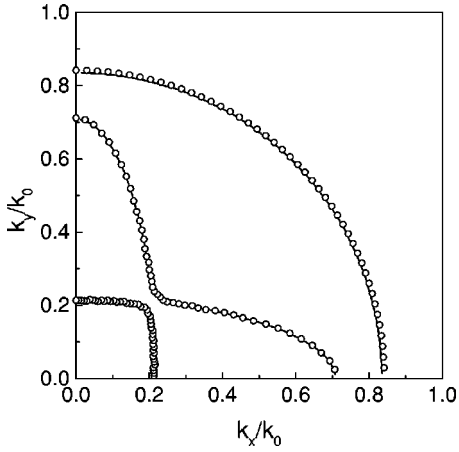


FIG. 6. Slowness surfaces for plane waves in an effective homogeneous medium (solid lines) and Floquet waves (open circle) in a periodic medium [0/90] in the $x-y$ plane. The slowness in this plane was not used for the effective elastic constants reconstruction and thus illustrates the equivalency of effective and periodic media.

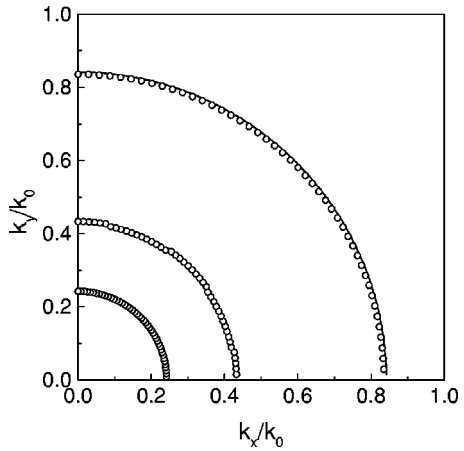


FIG. 7. Slowness surfaces for plane waves in a homogenized medium (solid lines) and Floquet waves (open circle) in a periodic medium [0/45/90/−45] in the $x-y$ plane. The slowness in this plane was not used for the effective elastic constants reconstruction and thus illustrates the equivalency of effective and periodic media.

increases and the other parameters depend slightly on frequency. This accords with the first critical angle decrease with frequency (Fig. 2), which is determined by the longitudinal velocity V_l propagating along the x (1) direction $V_l = \sqrt{C_{11}/\rho}$. This also agrees with the results of Datta[9] who used the effective stiffness model.

To check the accuracy of the effective constants, we calculated the slowness surfaces for Floquet waves and waves in the effective medium in the $x-y$ plane (the $x-y$ plane has not been used in the optimization procedure and the effective elastic constants reconstruction). Figures 6 and 7 show the slowness in the $x-y$ plane for Floquet waves (indicated by circles) and waves in the effective medium (indicated by solid line). As one can see they match very well.

To further validate our approach, we compare the effective elastic constants at very low frequency with static calculations of composite elastic constants using the generalized method of cells (GMC).[27] The cell [0/90] is used as an example. The frequency used to calculate effective elastic constants using the Floquet wave homogeneous method (FHM) is 100 Hz. The MAC software[28] was used to obtain the effective elastic constants using the GMC method. The results are shown in Table II. As can be seen, the results obtained by FHM are very close to those calculated by the static homogenization method.

B. Comparison of simulated time domain signals

As an additional illustration, let us compare the time domain signals transmitted through the multi-ply composite sample calculated in the plane wave approximation using the exact multilayered and homogenization theories. Figures 8 and 9 show this comparison for a pulse with center frequency at 1 MHz. The signals are simulated for a double-through-transmission through a laminate at various incident angles. The data for $[0/90]_{8s}$ composite are shown in Fig. 8 and for $[0/45/90/-45]_{16}$ composite in Fig. 9. The incident plane is oriented at 0°. Lamina properties of the composite are given in Table I. At normal incidence 0°, we observe multiple transmission of the quasilongitudinal wave. At 10°, both quasilongitudinal and quasishear waves appear. At large incident angles, only the quasishear wave is received. As shown in these figures, the signal calculated using effective elastic constants provides a good approximation of the signals calculated using exact theory. The phase and time delay of the first arrived wave, which is important for the velocity measurements in the double-through-transmission technique,[24,26] can be very accurately calculated using the effective elastic constants. This indicates that one can determine the effective elastic constants by the reconstruction method described in Refs. 24–26 from the measured time delays in the homogenization domain.

TABLE II. Results calculated by generalized cell method (GCM) and Floquet wave homogenization method (FHM) (unit: GPa).

	C_{11}	C_{33}	C_{12}	C_{23}	C_{44}	C_{66}
GCM	79.429	15.800	7.508	7.850	4.926	7.000
FHM	79.427	15.800	7.507	7.853	4.927	7.000

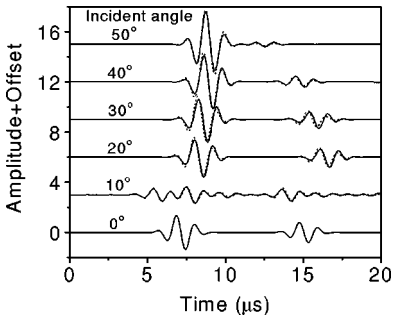


FIG. 8. Comparison of time domain signals calculated by exact and homogenization theory for $[0/90]_{32}$ composites at 1 MHz (double through transmission measurement). Solid lines are calculated by exact theory and dashed lines are calculated by homogenization theory.

IV. ESTIMATION OF THE HOMOGENIZATION DOMAIN FOR AN ANISOTROPIC CELL

As discussed above, the composite can be considered as an effective homogeneous medium at the frequency below the first stop band. To determine this frequency, one can use the Floquet wave spectrum zones calculated using Eq. (4) as

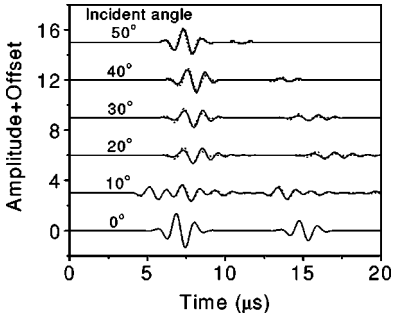


FIG. 9. Comparison of time domain signals calculated by exact and homogenization theory for $[0/45/90/-45]_{16}$ composites at 1 MHz (double through transmission measurement). Solid lines are calculated by exact theory and dashed lines are calculated by homogenization theory [(a) incident angle 0°, 10°, 20°; (b) incident angle 30°, 40°, 50°].

shown in Fig. 2. However, for applications it is useful to have a simple estimation method of the upper frequency bound of this domain. In this section, we will propose a procedure to estimate the homogenization domain using the velocities for each of the layers forming the cell.

Let us first consider a periodic medium whose cell has only two isotropic layers. Let us, for simplicity, consider SH wave propagation and assume that the velocities in the layers are different while the density and thickness ($h/2$) are the same for both layers. For such a system, the Floquet wave characteristic equation was obtained by Auld *et al.*,[21]

$$\cos(h\zeta)=\cos(\pi\alpha)\cos(\pi N\alpha) - \frac{1}{2}\left(N+\frac{1}{N}\right)\sin(\pi\alpha)\sin(\pi N\alpha), \tag{7}$$

where

$$\alpha=\frac{k_{z1}h}{2\pi} \quad \text{and} \quad N=\frac{k_{z2}}{k_{z1}}, \tag{8}$$

V_1 and V_2 are the velocities in layers one and two, respectively (we select $V_1<V_2$), ζ is the Floquet wave number. k_{z1} and k_{z2}, which are the z-direction components of the wave number in layers one and two, respectively, can be real or pure imaginary. If k_{z1} is imaginary and k_{z2} is real, we define $N'=\mathrm{Im}(k_{z2})/k_{z1}$. For this simple model, we can determine the homogenization zone upper bound as $\alpha<0.5$ and the Floquet wave critical angle as $N'=1$.

Below we generalize these criteria to our anisotropic case. If the cell is built from anisotropic laminas, the wave number k_z is not only dependent on incident angle (or wave number projection k_x) but also on incident plane orientation. Therefore the homogeneous domain can vary with incident angle and incident plane orientation. As follows from Eq. (8), the upper frequency bound of the homogenization domain ($\alpha<0.5$) depends only on the normal component of the wave number k_z and cell thickness h, i.e., on the ratio of the thickness to wavelength in the z direction. In anisotropic media, we have three plane waves. To find the condition equivalent to Eq. (8) we need to select the largest wave number $k_{z\,\max}$. The $k_{z\,\max}$ will depend on the propagation angle and orientation of the propagation plane.

$$\alpha=\frac{k_{z\,\max}(\theta)h}{2\pi}=\frac{h}{\lambda_{\min}(\theta)}\cos\theta<0.5 \quad \text{or} \quad f_h<\frac{0.25V_{\min}(\theta)}{h\cos\theta}, \tag{9}$$

where h is the cell thickness and $\lambda_{\min}$ and $V_{\min}=\omega/k_{\max}$ are the corresponding minimum wavelength and velocity at a given propagation angle. For example, for the cell [0/90], Fig. 10(a) shows the three wave number k_z components (slowness surface) for lamina 0° (solid lines) and for lamina 90° (dotted lines) for the $x-z$ incident plane. From Fig. 10(a) we can select the maximum $k_{z\,\max}$ among the six values for cell [0°/90°] and plot its value versus frequency for different incident angles θ_0 from fluid onto the laminated structure. Such a plot is given by the solid line in Fig. 10(b) separating shaded and unshaded areas. The homogenization domain is shown by the nonshaded area. The three critical angles are estimated from the condition $\mathrm{Im}(k_z^0)=k_z^{90}$ which is satisfied at the intersections marked as solid rectangles in Fig. 10(a). These three estimated critical angles are shown by the dotted lines in Fig. 10(b).

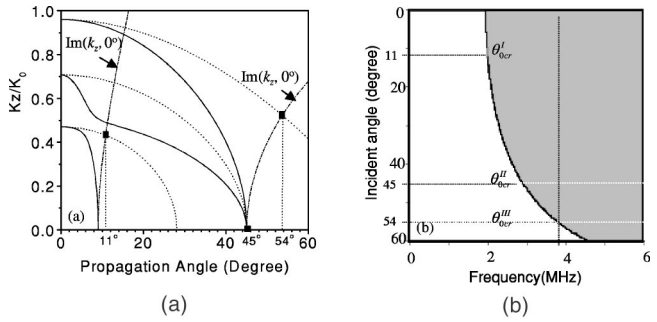


FIG. 10. Homogenization domain estimation for [0/90] composite cell. (a) Relations between k_z and incident angle (k_x) for 0° (solid line) and 90° (dotted lines) laminas. The dashed–dotted lines marked Im(k_z,0°) show the imaginary part of the nonhomogenous wave in the 0° lamina. The three solid rectangles indicate the intersections of the "Im(k_z,0°)" curves with the real part of the wave number in the 90° lamina. They determine the critical angles. (b) Predicted homogenization domain shown by unshaded area and three corresponding critical angles for [0/90] composite. The data are plotted for a wave incident from fluid with velocity 1500 m/s.

To illustrate the applicability of the results, we overlap these estimates of the homogenization domain with the Floquet wave pass and stop bands shown in Fig. 2. The estimated critical angles and homogenization zone are indicated by the dashed lines in Fig. 2. One can see that Eq. (9) gives a good estimate on the conservative side. Both estimation and exact theory show that the upper frequency bound of the homogenization domain increases as the incident angle increases.

V. CONCLUSION

Based on the slowness surface equivalence between a Floquet wave in a periodic medium and a plane wave in an effective homogeneous medium, a dynamic homogenization method has been proposed to determine the frequency-dependent elastic constants of the effective homogeneous medium. Computational examples for cross-ply and quasi-isotropic composites are given. The Floquet wave characteristic equation based on a computationally stable recursive stiffness matrix method is proposed. For [0/90] and [0/45/90/−45] composites, we identify a low frequency region, the homogenization domain, within which the periodic medium is equivalent to an effective homogeneous anisotropic medium. The upper frequency bound of this homogenization domain is estimated by a simple criterion, $(h/\lambda_{\min})\cos\theta<0.5$, where $\lambda_{\min}$ is the minimum wavelength in the cell layer at the given propagation direction θ. Numerical results show that time domain signals calculated using homogenization theory are very close to the exact results within the homogenization zone.

APPENDIX A: LAYER AND CELL STIFFNESS MATRICES

Let us consider an anisotropic cell $[\theta_1/\theta_2/\cdots\theta_{n-1}/\theta_n]$, consisting of n arbitrarily oriented generally anisotropic layers as illustrated in Fig. 1. In the mth layer, the displacement vector $\mathbf{u}^m$ may be written as the summation of six elastic waves,

$$\mathbf{u}^m=\sum_{j=1}^{3}\,(a_j^+\mathbf{p}_j^+e^{ik_z^{+j}(z-z_m)} + a_j^-\mathbf{p}_j^-e^{ik_z^{-j}(z-z_{m-1})})_m e^{i(k_x x+k_y y-\omega t)}, \tag{A1}$$

where $\mathbf{u}^m=(u_x^m,u_y^m,u_z^m)^T$, T represents transpose, j (=1,2,3) denotes jth wave. The superscripts + and − represent wave propagation in the $+z$ or $-z$ direction, respectively; $\mathbf{p}_j^+$ and $\mathbf{p}_j^-$ $[\mathbf{p}_j^\pm=(p_x^\pm,p_y^\pm,p_z^\pm)_j^T]$ are the unit displacement polarization vectors corresponding to waves with wave vector $(k_z^+)_j$ and $(k_z^-)_j$, respectively. The displacement polarization vectors $\mathbf{p}_j^\pm$ and wave numbers $k_z^{\pm j}$ are determined by solving the Christoffel equation[29]

$$(c_{ijln}k_jk_n-\rho\omega^2\delta_{il})p_l=0, \tag{A2}$$

where c_{ijln} represents the layer elastic constants and ρ the density.

Using Eqs. (A1) and (A2) and Hooke's law, one can obtain the layer transfer matrix $\mathbf{B}_m$ which relates stresses and displacements on the top $(z=z_{m-1})$ and bottom $(z=z_m)$ surfaces.

$$\begin{bmatrix}\mathbf{u}_m\\ \boldsymbol{\sigma}_m\end{bmatrix}=\begin{bmatrix}\mathbf{P}^-\mathbf{H}^- & \mathbf{P}^+\\ \mathbf{D}^-\mathbf{H}^- & \mathbf{D}^+\end{bmatrix}_m\begin{bmatrix}\mathbf{P}^- & \mathbf{P}^+\mathbf{H}^+\\ \mathbf{D}^- & \mathbf{D}^+\mathbf{H}^+\end{bmatrix}_m^{-1}\begin{bmatrix}\mathbf{u}_{m-1}\\ \boldsymbol{\sigma}_{m-1}\end{bmatrix}, \tag{A3}$$

where $\mathbf{P}^\pm(3\times3)=[\mathbf{p}_1^\pm,\mathbf{p}_2^\pm,\mathbf{p}_3^\pm]$, and $\mathbf{H}^+(3\times3)=\mathrm{Diag}[e^{ik_z^{+1}h_m},e^{ik_z^{+2}h_m},e^{ik_z^{+3}h_m}]$, $\mathbf{H}^-(3\times3)=\mathrm{Diag}[e^{-ik_z^{-1}h_m},e^{-ik_z^{-2}h_m},e^{-ik_z^{-3}h_m}]$, $h_m=z_{m-1}-z_m$ being the thickness of the mth lamina. If the z axis is a symmetry axis, we have $k_z^{-j}=-k_z^{+j}$, therefore $\mathbf{H}^+=\mathbf{H}^-$. This will be the case of monoclinic symmetry corresponding to arbitrary rotation of the orthotropic layer around the symmetry axis z. $\mathbf{D}^\pm=[\mathbf{d}_1^\pm,\mathbf{d}_2^\pm,\mathbf{d}_3^\pm]$. The components $(d_i^\pm)_j$ of $\mathbf{d}_j^\pm$ are related to the polarization vector $\mathbf{p}_j^\pm$ by $(d_i^\pm)_j=(c_{i3ln}k_np_l^\pm)_j$.

The stiffness matrix $\mathbf{K}_m$ relates stresses $(\boldsymbol{\sigma}_{m-1},\boldsymbol{\sigma}_m)$ to displacements $(\mathbf{u}_{m-1},\mathbf{u}_m)$ at layer top and bottom surfaces; rearranging correspondingly (A3), we have

$$\begin{bmatrix}\boldsymbol{\sigma}_{m-1}\\ \boldsymbol{\sigma}_m\end{bmatrix}=\begin{bmatrix}\mathbf{D}^- & \mathbf{D}^+\mathbf{H}^+\\ \mathbf{D}^-\mathbf{H}^- & \mathbf{D}^+\end{bmatrix}_m\begin{bmatrix}\mathbf{P}^- & \mathbf{P}^+\mathbf{H}^+\\ \mathbf{P}^-\mathbf{H}^- & \mathbf{P}^+\end{bmatrix}_m^{-1}\begin{bmatrix}\mathbf{u}_{m-1}\\ \mathbf{u}_m\end{bmatrix} =\begin{bmatrix}\mathbf{K}_m^{11} & \mathbf{K}_m^{12}\\ \mathbf{K}_m^{21} & \mathbf{K}_m^{22}\end{bmatrix}\begin{bmatrix}\mathbf{u}_{m-1}\\ \mathbf{u}_m\end{bmatrix}. \tag{A4}$$

The cell stiffness matrix is obtained from the layer stiffness matrices $\mathbf{K}^m$ by the recursive algorithm

$$\mathbf{K}^M=\begin{bmatrix}\mathbf{K}_{11}^{M-1}+\mathbf{K}_{12}^{M-1}(\mathbf{K}_{11}^m-\mathbf{K}_{22}^{M-1})^{-1}\mathbf{K}_{21}^{M-1} & -\mathbf{K}_{12}^{M-1}(\mathbf{K}_{11}^m-\mathbf{K}_{22}^{M-1})^{-1}\mathbf{K}_{12}^m\\ \mathbf{K}_{21}^m(\mathbf{K}_{11}^m-\mathbf{K}_{22}^{M-1})^{-1}\mathbf{K}_{21}^{M-1} & \mathbf{K}_{22}^m-\mathbf{K}_{21}^m(\mathbf{K}_{11}^m-\mathbf{K}_{22}^{M-1})^{-1}\mathbf{K}_{12}^m\end{bmatrix}, \tag{A5}$$

where $\mathbf{K}^M$ is the total stiffness matrix for the top m layers, $\mathbf{K}_{ij}^{M-1}$ are the submatrices of the total stiffness matrix for the top $m-1$ layers, $\mathbf{K}_{ij}^m$ are the stiffness submatrices for the mth layer.

APPENDIX B: EQUIVALENCE OF FLOQUET AND PLANE WAVES FOR A MATERIAL WITH HOMOGENIZED CELL

Here we will demonstrate that for an effective homogeneous anisotropic medium the Floquet wave wave numbers [calculated by Eq. (5)] are equal to the wave numbers of plane waves in the effective medium [calculated by Eq. (A2)]. Let us consider an infinite anisotropic homogeneous medium in which we select a layer (cell) and form a fictitious periodic medium. We use the transfer matrix $\mathbf{B}_c$ for this cell, which is in this case the transfer matrix for a single homogeneous anisotropic layer

$$\mathbf{B}_c=\begin{bmatrix}\mathbf{P}^-\mathbf{H}^- & \mathbf{P}^+\\ \mathbf{D}^-\mathbf{H}^- & \mathbf{D}^+\end{bmatrix}\begin{bmatrix}\mathbf{P}^- & \mathbf{P}^+\mathbf{H}^+\\ \mathbf{D}^- & \mathbf{D}^+\mathbf{H}^+\end{bmatrix}^{-1} =\begin{bmatrix}\mathbf{P}^- & \mathbf{P}^+\\ \mathbf{D}^- & \mathbf{D}^+\end{bmatrix}\begin{bmatrix}\mathbf{H}^- & 0\\ 0 & (\mathbf{H}^+)^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{P}^- & \mathbf{P}^+\\ \mathbf{D}^- & \mathbf{D}^+\end{bmatrix}^{-1}. \tag{B1}$$

From the periodicity condition (1) and the definition of the transfer matrix (A3) one obtains

$$\det\left(\begin{bmatrix}\mathbf{P}^- & \mathbf{P}^+\\ \mathbf{D}^- & \mathbf{D}^+\end{bmatrix}\begin{bmatrix}\mathbf{H}^- & 0\\ 0 & (\mathbf{H}^+)^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{P}^- & \mathbf{P}^+\\ \mathbf{D}^- & \mathbf{D}^+\end{bmatrix}^{-1}-e^{i\zeta h}\mathbf{I}\right)=0. \tag{B2}$$

Using the commutative property of the determinant of the matrix product we have

$$\det\left(\begin{bmatrix}\mathbf{H}^--e^{i\zeta h}\mathbf{I}_{3\times3} & 0\\ 0 & (\mathbf{H}^+)^{-1}-e^{i\zeta h}\mathbf{I}_{3\times3}\end{bmatrix}\right)=0, \tag{B3}$$

where $\mathbf{I}_{3\times3}$ is the 3×3 identity matrix and $\mathbf{H}^\pm(3\times3)=\mathrm{Diag}[e^{ik_z^{\pm1}h_m},e^{ik_z^{\pm2}h_m},e^{ik_z^{\pm3}h_m}]$ [see Eq. (A3)]. To satisfy equation (B3) the Floquet wave wave number ζ must be equal to one of the wave numbers k_z^i. This is applied to each of the three Floquet waves. Therefore the z-direction component of the wave numbers of Floquet waves and partial waves in an effective homogeneous medium are identical. The x-direction component will also be equal according to Snell's law. Thus for a homogeneous anisotropic medium,

the Floquet wave velocities equal the effective wave velocities. This was also noted by Nayfeh (Ref. 12, p. 130).

[1] J. D. Achenbach, C. T. Sun, and G. Herrmann, J. Appl. Mech. **35**, 689–696 (1968).
[2] C. T. Sun, J. D. Achenbach, and G. Herrmann, J. Appl. Mech. **35**, 408–411 (1968).
[3] C. T. Sun, J. D. Achenbach, and G. Herrmann, J. Appl. Mech. **35**, 467–475 (1968).
[4] J. D. Achenbach, ''Wave propagation in fiber reinforced composites,'' in *Composite Materials, Vol. 2, Mechanics of Composite Materials*, edited by G. P. Sendeckyi (Academic, New York, London, 1974), Chap. 8.
[5] A. H. Shah and S. K. Datta, Int. J. Solids Struct. **18**, 297–410 (1982).
[6] S. K. Datta, A. H. Shah, and H. M. Ledbetter, in *Mechanics of Composite Materials*, edited by Z. Hashin and C. T. Herakovich (Pergamon, New York, 1983).
[7] W. Karunasena, A. H. Shah, and S. K. Datta, J. Appl. Mech. **58**, 1028–1032 (1991).
[8] S. K. Datta, T. H. Ju, R. L. Bratton, and A. H. Shah, Int. J. Solids Struct. **29**, 1711–1721 (1992).
[9] S. K. Datta, in *Comprehensive Composite Materials*, edited by T. W. Chou (Elsevier, New York, 2000), Chap. 1.18, pp. 511–558.
[10] *Wave Propagation in Structural Composites*, edited by A. K. Mal and T. C. T. Ting (ASME–AMD-Vol. 90, ASME, New York, NY, 1988).
[11] A. H. Nayfeh, J. Acoust. Soc. Am. **89**, 1521–1528 (1991).
[12] A. H. Nayfeh, *Wave Propagation in Layered Anisotropic Media* (North-Holland, Amsterdam, 1995).
[13] D. E. Chimenti and A. H. Nayfeh, J. Acoust. Soc. Am. **87**, 1409–1415 (1990).
[14] H. Nayfeh and D. E. Chimenti, J. Acoust. Soc. Am. **89**, 542–554 (1991).
[15] L. Wang and S. I. Rokhlin, Ultrasonics **39**, 413–424 (2001).
[16] A. M. Braga and G. Herrmann, J. Acoust. Soc. Am. **91**, 1211–1221 (1992).
[17] C. Potel, J. Belleval, and Y. Gargouri, J. Acoust. Soc. Am. **97**, 2815–2826 (1995).
[18] M. Schoenberg, Wave Motion **6**, 302–320 (1984).
[19] B. Hosten, Ultrasonics **30**, 365–371 (1992).
[20] P. J. Shull, D. E. Chimeti, and S. K. Datta, J. Acoust. Soc. Am. **95**, 99–105 (1994).
[21] A. Auld, D. E. Chimenti, and P. J. Shull, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 319–326 (1996).
[22] S. I. Rokhlin, X. Qiang, Y. Liu, and L. Wang, in *Review of Progress in QNDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1999), Vol. 18B, pp. 1249–1256.
[23] L. Wang and S. I. Rokhlin, in *Review of Progress in QNDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1999), Vol. 18B, pp. 1321–1328.
[24] S. I. Rokhlin and W. Wang, J. Acoust. Soc. Am. **91**, 3303–3312 (1992).
[25] Y. C. Chu and S. I. Rokhlin, J. Acoust. Soc. Am. **95**, 213–225 (1994).
[26] Y. C. Chu, A. D. Degtyar, and S. I. Rokhlin, J. Acoust. Soc. Am. **95**, 3191–3203 (1994).
[27] M. Paley and J. Aboudi, Mech. Mater. **14**, 127–140 (1992).
[28] S. M. Arnold, B. A. Bednarcyk, T. E. Wilt, and D. Trowbridge, Micromechanics analysis code with generalized method of cells, NASA/TM 1999-209070, March, 1999.
[29] S. I. Rokhlin, T. K. Bolland, and L. Adler, J. Acoust. Soc. Am. **79**, 906–910 (1986).

On the effect of viscosity in scattering from partially coated infinite cylinders

Jerry H. Ginsberg[a)]
G. W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 4 January 2002; accepted for publication 28 March 2002)

This paper considers the two-dimensional problem of scattering of a plane wave incident on an infinite cylinder that is coated with strips of pressure-release material extending over quadrants on the illuminated and shadowed sides, with the remainder of the surface considered to be rigid. Transitions from soft to rigid surfaces correspond to discontinuous boundary conditions. Ideal fluid theory predicts an infinite pressure gradient at these transitions, which suggests that viscous effects may be significant. The present work is a quantitative analysis of the global effect on acoustic scattering of viscosity effects arising in the vicinity of the discontinuity. The analysis represents the scattered field in terms of acoustic and vortical contributions. Both contributions are represented by series expansions in terms of azimuthal harmonics and associated cylindrical wave functions. The amplitudes of these harmonics are determined by satisfying a pair of discontinuous boundary conditions. Results obtained by using the method of weighted residuals are shown to be less accurate than those obtained from a collocation procedure. The results for surface pressure and farfield directivity indicate that viscous effects are important only if the Reynolds number is extremely small. [DOI: 10.1121/1.1479147]

PACS numbers: 43.20.Fn, 43.20.Tb, 43.35.Mr [ADP]

I. INTRODUCTION

It is common practice to cover a submerged structure with a rubberlike substance in order to alter its acoustical performance. Doing so over only a portion of the wetted surface might enable one to obtain the desired acoustical properties more economically. However, a corollary of partial coating coverage is the creation of discontinuities in surface impedance whose effect on scattering of acoustic waves has not been thoroughly explored. One such effect is immediately apparent: since the pressure must vary continuously across an edge separating regions having different impedance, the normal velocity must vary discontinuously across that edge. Ferri, Ginsberg, and Rogers[1] considered acoustic scattering from a cylinder which is partially coated with regions of a soft material, represented as a pressure–release coating, separated by uncoated regions that were considered to be rigid. They showed that the normal velocity on the coated side depends inversely on the square root of the distance from the edge. Correspondingly, the pressure on the uncoated side approaches an infinite tangential gradient with decreasing distance from the edge. These singular spatial variations are confined to a small region around the edge whose extent is essentially independent of frequency, and therefore of the compressibility of the fluid. In a real fluid, steep pressure gradients are accompanied by large shear stresses. The present investigation examines the degree to which viscosity effects that are strongest in the vicinity of the discontinuity globally affect the acoustic scattering process.

Several investigators had addressed scattering phenomena from partially coated cylinders and spheres prior to the work of Ferri *et al.* Babailov[2] studied axisymmetric scattering from a sphere with a coating covering the polar cap at which a plane wave is incident. The study explored the nature of the discontinuity in the surface response, but the global analysis was carried out by imposing conditions at the edge of the cap that are inconsistent with the assumption that the fluid is ideal. The investigation by Büyükaksoy and Uzgören[3] considered the two-dimensional situation where a coating patch is applied to the incident side of an infinitely long cylinder. Their analysis employed techniques that are only suitable for extremely high frequencies, such as considering the circumference on both sides of the patch to extend to infinity. Consequently, the work is more relevant to issues pertaining to electromagnetic scattering. Shaw[4,5] employed the CHIEF boundary element program to study the signal backscattered from a cylinder that is fully coated, except for small patches where the surface is rigid. The goal of that study was to determine the relationship between the backscattered farfield and the spatial extent and distribution of uncoated regions. Feit and Cruschieri[6] used a weighted residuals formulation to study radiation and scattering from a partially coated infinite cylinder. The structural model for their study was a shell, with the coating treated as a locally reacting thin layer. None of these works attempted to identify the role of nonideal effects in the fluid.

Related studies of scattering from surfaces containing impedance discontinuities have addressed the case where the wetted surface is the planar boundary of an ideal fluid in an infinite half-space. Works by Naghieh and Hayek,[7] Habault and Filippi,[8] Dahl and Frisk,[9] Tjøtta,[10] and others, employed a variety of analytical procedures, including the Weiner–Hopf transformation, that apply when the scattered wave consists of a continuous spectrum of wave numbers. In contrast, an infinite cylinder represents a separable geometry for the Helmholtz equation, so the scattered field may be repre-

a)Electronic mail: jerry.ginsberg@me.gatech.edu

sented by an infinite series corresponding to a discrete spectrum of wave numbers.

Several investigators have addressed the effects of viscosity on scattering from spheres and infinite cylinders whose surface impedance is uniform. The relevant research was surveyed in the paper by Hasheminejad and Geers,[11] who explored mutual scattering effects between two spheres. The basic concept underlying studies of uniformly coated bodies is the ability to decompose the scattered signal into acoustic, vorticity, and entropy modes, each of which is governed by a Helmholtz equation having a complex wave number. One would anticipate that thermoviscous effects become increasingly important with decreasing Reynolds numbers, $\mathrm{Re}=\rho ca/\mu$, where ρ, c, and μ are, respectively, the density, sound speed, and shear viscosity at ambient conditions, and a is the radius. An important conclusion reached by Hasheminejad and Geers was that even for small radii corresponding to bubbles, viscous effects usually are much more significant than thermal conduction. Another conclusion was that the scattered pressure is essentially unaffected by viscosity unless the radius is very small, as in the case of microbubbles.

The present study is concerned with relatively large diameter cylinders, corresponding to large Reynolds numbers. Accordingly, it is assumed from the outset that one can ignore thermal waves. Because ideal fluid theory leads to a surface response that contains singularities along edges of impedance discontinuity, one cannot say with certainty that viscous effects are unimportant. The basic analysis employs the field equations derived by Pierce.[12] It is assumed that the process is isentropic, and that temperature fluctuations relative to ambient conditions are sufficiently small to permit considering viscosity to be constant. One might think that nonlinear terms could also be important, but the velocity singularity is an artifice of the ideal fluid model. The viscous model will be seen to lead to pressure and particle velocities that are not extraordinarily large anywhere. Thus, nonlinear effects may be neglected, subject to the usual limitations on signal levels.

An auxiliary aspect of the following work pertains to the general issue of treating discontinuous boundary conditions. As the foregoing survey indicates, weighted residual formulation based on a global series representation and collocation-type formulation that satisfy the field equations at selected points have been used in the past. The work that follows provides some general insights to the merits of each approach.

II. FORMULATION

The basic equations are those derived by Pierce,[12] from which an $\exp(-i\omega t)$ time dependence is factored out. The total pressure and particle velocity are the sum of the incident and scattered terms. The dimensional pressure and particle velocity associated with the scattered signal are $\rho c^2 p$ and $c\bar{v}$, respectively; the incident signal is denoted in a similar manner with the addition of subscript "*I*." Because the incident signal is itself a solution of the field equations, the scattered signal must separately satisfy those equations. Let ar measure radial distance, and let ∇ be the corresponding nondimensional gradient. The linearized Navier–Stokes equations are

$$\nabla\cdot\bar{v}=ikap,$$
$$\nabla p-\frac{1}{\mathrm{Re}}\left[\nabla^2\bar{v}+\frac{1}{3}\nabla(\nabla\cdot\bar{v})\right]=ika\bar{v}. \tag{1}$$

The velocity is decomposed into potential and solenoidal components,

$$\bar{v}=\nabla\phi+\nabla\times\bar{\psi},\nabla\cdot\bar{\psi}=0. \tag{2}$$

Uncoupled equations for p and $\bar{\psi}$ are obtained by substituting Eq. (2) into Eqs. (1), and then taking the divergence and curl of the second equation. The result is

$$\nabla\cdot\bar{v}=\nabla^2\phi=ikap, \tag{3a}$$

$$\left(1-\frac{4ika}{3\,\mathrm{Re}}\right)\nabla^2 p+k^2a^2p=0, \tag{3b}$$

$$\nabla^2\bar{\psi}+ika\,\mathrm{Re}\,\bar{\psi}=\bar{0}. \tag{3c}$$

A simple relation for the velocity corresponding to solutions for p and $\bar{\psi}$ may be obtained by using Eq. (3b) to eliminate p in favor of $\nabla^2 p$ in Eq. (3a), which when integrated yields

$$\phi=\left(\frac{1}{ika}-\frac{4}{3\,\mathrm{Re}}\right)p. \tag{4}$$

It follows from Eq. (2) that the velocity is given by

$$\bar{v}=\left(\frac{1}{ika}-\frac{4}{3\,\mathrm{Re}}\right)\nabla p+\nabla\times\bar{\psi}. \tag{5}$$

The solutions for p and $\bar{\psi}$ in combination with the incident signal must satisfy boundary conditions on the surface of the cylinder. The total pressure at the boundary is assumed to be proportional to the normal (outward) velocity through a local specific impedance z that may vary with circumferential position, so that

$$p+\rho cz\bar{n}\cdot\bar{v}=-p_I-\rho cz\bar{n}\cdot\bar{v}_I \qquad \text{at } r=1. \tag{6}$$

The contact condition requires that there be no net velocity in the tangential direction, so that

$$\bar{n}\times\bar{v}=-\bar{n}\times\bar{v}_I \qquad \text{at } r=1. \tag{7}$$

The present two-dimensional problem concerns scattering from a rigid cylinder. In terms of polar coordinates centered on the cylinder, p and $\bar{\psi}=\psi\bar{k}$ may depend on r and θ only, in which case the requirement that $\nabla\cdot\bar{\psi}=0$ is satisfied identically. The corresponding velocity components obtained from Eq. (5) are

$$v_r=\left(\frac{1}{ika}-\frac{4}{3\,\mathrm{Re}}\right)\frac{\partial p}{\partial r}+\frac{1}{r}\frac{\partial\psi}{\partial\theta},$$
$$v_\theta=\left(\frac{1}{ika}-\frac{4}{3\,\mathrm{Re}}\right)\frac{1}{r}\frac{\partial p}{\partial\theta}-\frac{\partial\psi}{\partial r}. \tag{8}$$

The incident signal is a plane wave propagating in the negative x direction, in which case the scattered signal is

symmetric with respect to $\theta=0$. Suitable Fourier series expansions are therefore

$$p=\sum_{n=0} f_n(r)\cos(n\theta), \quad \psi=\sum_{n=1} g_n(r)\sin(n\theta). \tag{9}$$

These expansions satisfy Eqs. (3b) and (3c) if

$$\begin{aligned} &f_n''+\frac{1}{r}f_n'+\left(\lambda^2-\frac{n^2}{r^2}\right)f_n=0,\\ &g_n''+\frac{1}{r}g_n'+\left(i\eta-\frac{n^2}{r^2}\right)g_n=0, \end{aligned} \tag{10}$$

where

$$\begin{aligned} &\lambda=ka\Big/\left(1-\frac{4ika}{3\,\mathrm{Re}}\right)^{1/2}\approx ka\left(1+\frac{2ika}{3\,\mathrm{Re}}\right),\\ &\eta=(ka\,\mathrm{Re})^{1/2}. \end{aligned} \tag{11}$$

Note that the branch cuts associated with the definitions of λ and η are selected to assure that the ensuing solution satisfies the Sommerfeld radiation condition associated with the $\exp(-i\omega t)$ convention. These solutions are

$$f_n=A_nH_n(\lambda r), \quad g_n=B_nH_n(e^{i\pi/4}\eta r), \tag{12}$$

where H_n is an abbreviated notation for the Hankel function of the first kind. The corresponding scattered pressure and velocity components obtained by substituting Eqs. (9) into Eqs. (8) are

$$\begin{aligned} v_r=\sum_{n=0}\Bigg[&\left(\frac{1}{ika}-\frac{4}{\mathrm{Re}}\right)\lambda A_nH_n'(\lambda r)\\ &+\frac{n}{r}B_nH_n(e^{i\pi/4}\eta r)\Bigg]\cos(n\theta), \end{aligned} \tag{13a}$$

$$\begin{aligned} v_\theta=-\sum_{n=1}\Bigg[&\left(\frac{1}{ika}-\frac{4}{3\,\mathrm{Re}}\right)\frac{n}{r}A_nH_n(\lambda r)\\ &+e^{i\pi/4}\eta B_nH_n'(e^{i\pi/4}\eta r)\Bigg]\sin(n\theta), \end{aligned} \tag{13b}$$

$$p=\sum_{n=0} A_nH_n(\lambda r)\cos(n\theta). \tag{13c}$$

Upon substitution of Eqs. (13a)–(13c), the boundary conditions, Eqs. (6) and (7), require that

$$\begin{aligned} \sum_{n=0}^{\infty}\Bigg\{&\left[H_n(\lambda)+z\left(\frac{1}{ika}-\frac{4}{3\,\mathrm{Re}}\right)\lambda H_n'(\lambda)\right]A_n\\ &+[znH_n(e^{i\pi/4}\eta)]B_n\Bigg\}\cos(n\theta)=-[P_I+z(v_I)_r]\big|_{r=1}, \end{aligned} \tag{14a}$$

$$\begin{aligned} \sum_{n=1}^{\infty}\Bigg\{&\left[\left(\frac{1}{ika}-\frac{4}{3\,\mathrm{Re}}\right)nH_n(\lambda)\right]A_n\\ &+[e^{i\pi/4}\eta H_n'(e^{i\pi/4}\eta)]B_n\Bigg\}\sin(n\theta)=(v_I)_\theta\big|_{r=1}. \end{aligned} \tag{14b}$$

III. WEIGHTED RESIDUALS SOLUTION

In the case of a uniform coating (constant z), Eqs. (14a) and (14b) are readily satisfied by invoking orthogonality of the azimuthal harmonics. A comparable solution for the case where z is a function of θ is the method of weighted residuals. This entails truncating the Fourier series expansions, Eq. (9), at a sufficiently large harmonic N. Multiplication of Eq. (14a) by $\cos(j\theta)$ for $j=0,...,N$ and Eq. (14b) by $\sin(j\theta)$ for $j=1,...,N$, followed by integration of each over $0<\theta<\pi$, yields $2N+1$ linear equations for the unknown coefficients. The result is a set of simultaneous equations, whose partitioned form is

$$\begin{bmatrix}[N]^p & [N]^\psi\\ [T]^p & [T]^\psi\end{bmatrix}\begin{Bmatrix}A\\ B\end{Bmatrix}=-\begin{Bmatrix}P+W\\ V\end{Bmatrix}. \tag{15}$$

In the foregoing, the coefficient matrices $[N]^p$ and $[N]^\psi$ are associated with the normal boundary condition, Eq. (14a), while $[T]^p$ and $[T]^\psi$ are associated with the tangential boundary condition, Eq. (14b). The coefficients on the right-hand side represent the excitations of the incident wave. The system matrix coefficients are

$$\begin{aligned} &(N_{jn})^p=\frac{\pi}{2}H_j(\lambda)\varepsilon_j\delta_{jn}+\left(\frac{1}{ika}-\frac{4}{3\,\mathrm{Re}}\right)\lambda H_n'(\lambda)C_{nj},\\ &(N_{jn})^\psi=nH_n(e^{i\pi/4}\eta)C_{nj},\\ &(T_{jn})^p=\frac{\pi}{2}\left(\frac{1}{ika}-\frac{4}{3\,\mathrm{Re}}\right)jH_j(\lambda)\delta_{jn},\\ &(T_{jn})^\psi=\frac{\pi}{2}e^{i\pi/4}\eta H_j'(e^{i\pi/4}\eta)\delta_{jn}, \end{aligned} \tag{16}$$

where $\varepsilon_j=2$ if $j=0$, and $\varepsilon_j=1$ otherwise, and δ_{jn} is the Kronecker delta. The coefficients C_{nj} represent integrals that depend on $z(\theta)$, while P_j, W_j, and V_j contain the effects of the incident wave,

$$\begin{aligned} &C_{nj}=\int_0^\pi z\cos(n\theta)\sin(j\theta)\,d\theta,\\ &P_j=\int_0^\pi P_1\cos(j\theta)\,d\theta,\\ &W_j=\int_0^\pi z(v_I)_r\cos(j\theta)\,d\theta,\\ &V_j=\int_0^\pi (v_I)_\theta\sin(j\theta)\,d\theta. \end{aligned} \tag{17}$$

Note that because $[T]^p$ and $[T]^\psi$ are both diagonal, it is a simple matter to eliminate $\{B\}$ from Eqs. (15), thereby reducing the problem to solution of $N+1$ equations.

The incident wave is planar, traveling in the negative x direction, from which it follows that

$$\begin{aligned} &P_I=P_0\exp(-ika\cos\theta),\\ &(v_I)_r=-P_0\exp(-ika\cos\theta)\cos(\theta),\\ &(v_I)_\theta=P_0\exp(-ika\cos\theta)\sin(\theta). \end{aligned} \tag{18}$$

In the situation treated, here z is a piecewise constant function of θ, which permits analytical evaluation of P_j and V_j in Eqs. (17). An alternative to invoking a numerical algorithm to evaluate W_j is to employ a Fourier series representation of the impedance. Truncation of such a series at the highest harmonic used to represent the scattered signal leads to

$$z=\sum_{n=0}^{N} Z_j \cos(n\theta). \tag{19}$$

The resulting coefficients are

$$\begin{aligned} P_j &= P_0 \pi (-i)^j J_j(ka), \\ V_j &= \tfrac{1}{2} P_0 \pi [(-i)^{|j-1|} J_{|j-1|}(ka) - (-i)^{(j+1)} J_{(j+1)}(ka)], \end{aligned} \tag{20}$$

$$\begin{aligned} W_j = -\frac{\pi}{4} P_0 \sum_{n=0}^{N} Z_n \{ & (-i)^{|j-n+1|} J_{|j-n+1|}(ka) \\ & + (-i)^{|j+n-1|} J_{|j+n-1|}(ka) \\ & + (-i)^{|j-n-1|} J_{|j-n-1|}(ka) \\ & + (-i)^{(j+n+1)} J_{(j+n+1)}(ka) \}. \end{aligned}$$

Note that as a consequence of the representation of impedance as a Fourier series, the highest order Bessel function requiring evaluation is $2N+1$. Several values of W_j obtained from this procedure were verified to be in extremely close agreement with the results of numerical integration.

IV. COLLOCATION SOLUTION

The difficulty with the method of weighted residuals is the weak nature of the solution, which merely asserts that the error in satisfying the boundary conditions is orthogonal to the N harmonics used to represent the scattered signal. The example will demonstrate that this approach can lead to significant errors in satisfying the boundary conditions locally. An alternative strategy employs the collocation method to satisfy the boundary conditions at a set of discrete points. This was the method employed previously by Ferri *et al.*[1] to investigate the analogous inviscid problem.

The first consideration in applying the collocation method is selection of the collocation points. Truncation of the series for pressure and vorticity at the Nth azimuthal harmonic leads to $2N+1$ unknown harmonic amplitudes. This poses a dilemma, because satisfying the boundary condition at M points yields $2M$ collocation equations. The resolution of this difficulty lies in the recognition that the boundary condition for tangential velocity is identically satisfied at $\theta=0,\ \pi$. Selecting either of these points will yield an odd number of collocation equations. The earlier analysis of the inviscid problem disclosed that it is best to employ equally spaced collocation points. Hence, letting the incident point, $\theta=0$, be the first point leads to enforcement of the boundary conditions at the $N+1$ points $\theta_n=(n-1)\Delta$, where $\Delta=2\pi/(2N+1)$. (Evaluations using $\theta=\pi$ as the last collocation point, rather than $\theta=0$ as the first, were found to yield results that are virtually identical to those obtained from the foregoing scheme.) Enforcing the boundary conditions, Eqs. (14a) and (14b) at the collocation points yields $2N+1$ simultaneous equations,

$$\begin{bmatrix} [\hat{N}]^p & [\hat{N}]^\psi \\ [\hat{T}]^p & [\hat{T}]^\psi \end{bmatrix} \begin{Bmatrix} A \\ B \end{Bmatrix} = -\begin{Bmatrix} \hat{P}+\hat{W} \\ \hat{V} \end{Bmatrix}, \tag{21}$$

where the various matrix elements are the respective coefficients in Eqs. (14a) and (14b) evaluated at the collocation points. Note that in this formulation, each partition of the system matrix is fully populated, so one is left to solve $2N+1$ equations.

A crucial aspect of the earlier work by Ferri *et al.* pertains to the manner in which the solution for the amplitudes is used to construct the surface response. The correct approach entails using the series representations to synthesize the scattered surface pressure and particle velocity at the collocation points only, with values between those points obtained by interpolation. The alternative of using the series elsewhere than the collocation point leads to significant numerical noise, as a corollary of the Nyquist criteria for the problem.

V. COMPUTATIONAL CONSIDERATIONS

Except for situations where the radius a is extremely small, the Reynolds number is quite large. The system coefficients multiplying $\{B\}$ in both Eqs. (15) and (21) require evaluation of $H_j(e^{i\pi/4}\eta)$ and its first derivative. In view of the definition of η, Eq. (11), the argument of these Hankel function has large imaginary and real parts. Furthermore, the order of these functions can be quite large, depending on the value of N. The generation of Bessel and Hankel functions having complex argument and high integer order has been the subject of several papers. For example, Mason[14] and Du Toit[13] give different algorithms. The present investigation uses a set of algorithms that are like Mason's, but somewhat simpler and more efficient. In any event, one will encounter difficulty with the Hankel function when η is large, as is evident from the principal asymptotic form (Abramowitz and Stegun[15]),

$$\begin{aligned} H_j(e^{i\pi/4}\eta) \approx e^{-i\pi/8} \left(\frac{2}{\pi\eta}\right)^{1/2} \exp\left(\frac{\eta}{\sqrt{2}}\right) \\ \times \exp\left[i\left(\frac{\eta}{\sqrt{2}} - j\pi/2 - \pi/4\right)\right]. \end{aligned} \tag{22}$$

A computational underflow occurs when η is very large (>1000 for double precision in the computer used for the present study), resulting in vanishing of the coefficients of B_j in some or all of the equations to be solved. This condition may be circumvented by scaling the coefficients B_j, such that $(B_j)_{\text{scaled}}=B_j H_j(e^{i\pi/4}\eta)$. When the system equations are recast accordingly, it is only necessary to compute the ratio $H_j'(e^{i\pi/4}\eta)/H_j(e^{i\pi/4}\eta)$. If η exceeds the computable range, one may employ the approximation $H_j'(e^{i\pi/4}\eta)/H_j(e^{i\pi/4}\eta)\approx i$. This is valid if the argument is large compared to j, which is true for the present analysis where evaluations are performed for N up to 200.

Evaluation of the Bessel functions $H_j(\lambda)$ and $J_j(ka)$, which arise in various terms of the weighted residual and collocation solutions, leads to a different problem. The difficulty is associated with the fact that the order of the Bessel function may be much larger than the argument. The corresponding function values may be extraordinarily large, which leads in some cases to ill-conditioning of the simultaneous equations. Once again, this situation may be handled by introducing scaled coefficients. The computation of the Bessel functions always begins by evaluating the values of $J_j(x)$ using a downward recursion algorithm. At each step in this process, the functional values are normalized to have a magnitude less than unity, and the corresponding power of 10 is stored separately. The next step is to use upward recursion to compute the Neumann functions $Y_j(x)$, whose upward recursion is stable if the imaginary part of the argument is small. At each step, the Neumann function values are scaled down by the factor used to scale the Bessel functions. The scaled Bessel and Neumann functions are then combined to form scaled Hankel functions. Correspondingly, the coefficients A_j in Eqs. (15) and (21) are scaled upward by the power of 10 used to scale down the Hankel functions.

An additional benefit of scaling both amplitudes in the aforementioned manner becomes evident when one considers the series expansions in Eqs. (13a)–(13c). Because the scaled Hankel functions have an order of magnitude that is close to unity, the magnitude of A_j and B_j indicate directly the contribution of the jth harmonic of p and ψ, respectively.

VI. EVALUATIONS

A useful check is to evaluate the case of a uniform coating. Correspondingly, the weighted residual equations uncouple, and the amplitudes reduce to those obtained from a conventional separation of variables analysis. Viscosity effects are negligible in this case, unless a is very small, see Hasheminejad and Geers. This is evident from Eqs. (15), where the second set of partitioned equations indicates that the order of magnitude of B_j is $1/\eta$ times the magnitude of A_j. For that reason results were computed for an unrealistic situation, Re=300, which corresponds in water to radius $a \approx 20$ nm. Figure 1 compares the (scaled) amplitudes obtained from the weighted residual and collocation formulations when $z=0.2$. Recall that the equations for the latter analysis are always fully coupled, regardless of the circumferential variation of surface impedance. It is apparent from Fig. 1 that both formulations yield the same surface response when the impedance is uniform.

The remainder of the discussion shall address the configuration studied by Ferri *et al.*[1] The incident and shadowed sides of the cylinder, extending from the plane of symmetry to $\pi/4$ are considered to be covered by a soft coating, while the top and bottom are considered to be a hard uncoated surface. Thus, $z(\theta)=z_1$ if $0 \leqslant |\theta| < \pi/4$ or $3\pi/4 < |\theta| \leqslant \pi$, while $z(\theta)=z_2$ if $\pi/4 < |\theta| < 3\pi/4$. Results reported here are for $z_1=0$ and $z_2=100$, corresponding to a pressure release coating on an otherwise rigid cylinder. [Equations (15) become ill-conditioned if one sets z_2 too large.]

The first issue to be addressed is convergence of each formulation. Once again, the analysis considers the value,

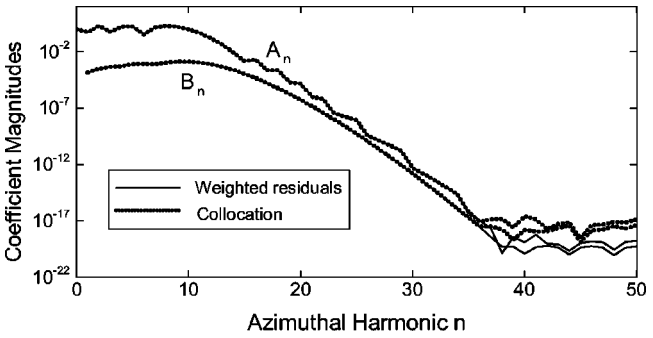


FIG. 1. Dependence of the pressure series coefficients in the azimuthal harmonic number for a uniformly coated cylinder according to weighted residuals and collocation, $z_1 = z_2 = 0.2$, Re=300, $ka=10$, $N=200$.

Re=300, in order to exaggerate the effect of viscosity. The frequency is set at $ka=10$. Figure 2 shows the values of the A_n coefficients obtained from weighted residuals, while Fig. 3 shows the corresponding data for the B_n coefficients. It is important to recall that the Hankel functions associated with each set of coefficients are scaled to have unit order of magnitude, so the coefficient values are a direct measure of the contributions of the respective harmonics. The results obtained from $N=25$ differ significantly from those for the higher N, but none seem to have converged. The highest order terms for each series length increase in magnitude with increasing n, but the lower order coefficients obtained from different N are in good agreement. For example, the amplitudes obtained for $N=100$ and $N=200$ diverge beyond $j=70$. Another interesting observation is that for each series length other than $N=25$, the coefficients for the last 25 harmonics are the ones that tend to increase.

The collocation solution for the same parameters is shown in Figs. 4 and 5. Clearly, the convergence properties of this approach are superior to those of the weighted residual solution. Truncations at $N=100$ and 200 yield very similar values for all amplitudes, up to $j=100$, and there is

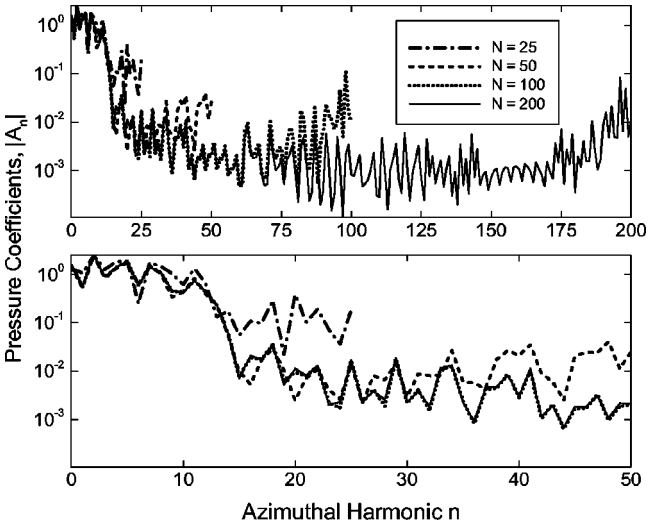


FIG. 2. Dependence of the pressure series coefficients in the azimuthal harmonic number for a partially coated cylinder according to weighted residuals, $z_1=0$, $z_2=100$, Re=300, $ka=10$.

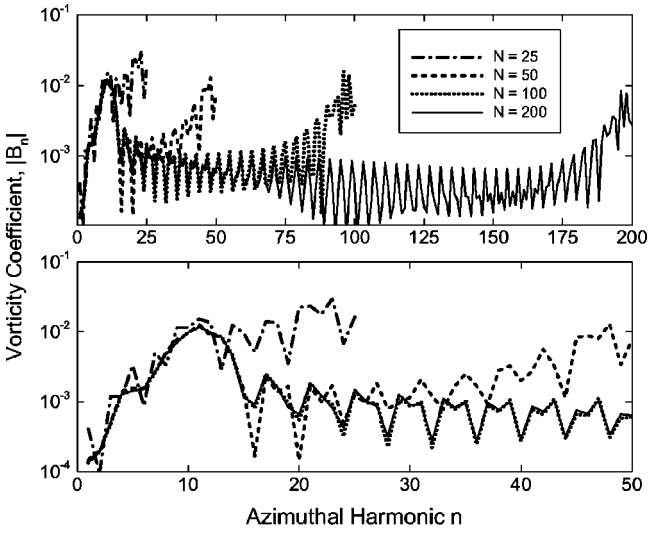


FIG. 3. Dependence of the vorticity series coefficients on the azimuthal harmonic number for a partially coated cylinder according to weighted residuals, $z_1=0$, $z_2=100$, Re=300, $ka=10$.

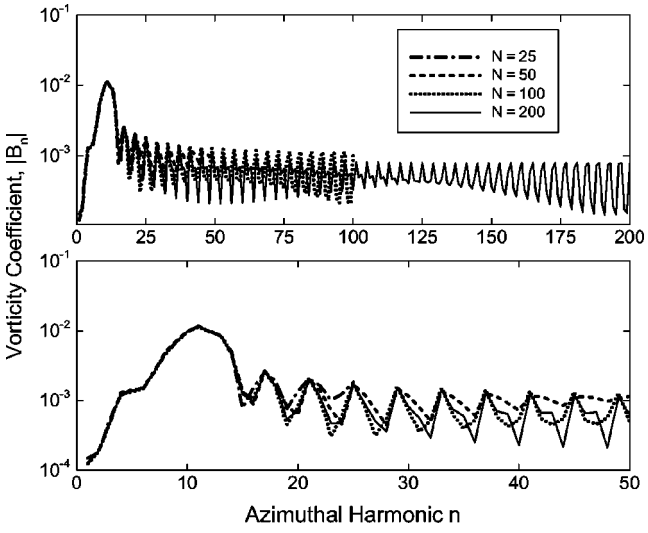


FIG. 5. Dependence of the vorticity series coefficients on the azimuthal harmonic number for a partially coated cylinder according to collocation, $z_1=0$, $z_2=100$, Re=300, $ka=10$.

no general tendency to increase at the highest harmonic numbers.

The corresponding pressure and radial velocity distributions on the surface obtained by evaluating Eqs. (13c) and (13a) at the collocation points appear in Figs. 6 and 7. The series length for each evaluation is $N=200$. The largest gradients of pressure and radial velocity occur, respectively, on the uncoated and coated sides of each impedance discontinuity. It is evident from these figures that the solutions obtained from weighted residuals and collocation are in overall agreement, but the former is somewhat in error, to the extent that it does not precisely satisfy the boundary condition that $p=0$ for $0<\theta<45°$ and $135°<\theta<180°$. The discrepancy between the two solutions stems from the fact that the boundary condition is only enforced in an average sense in weighted residuals. Interestingly, the results obtained from both techniques yield values for v_r that are very small over $45°<\theta<135°$, as required by the fact that z is large in that region.

Figures 6 and 7 also display the solution obtained when viscosity is ignored, Re$=\infty$. Of course, a collocation analysis cannot actually evaluate the singular radial velocity at the discontinuity, but the tendency is evident. Overall, the basic features of both the pressure and radial velocity distributions are quite similar, despite the very low value of Re. Other than reducing the radial velocity in the immediate vicinity of the discontinuities, viscosity has little effect on the magnitude of the surface response. However, it is interesting to observe that the pressure obtained with viscosity grows slightly more rapidly beyond $\theta=45°$, resulting in advancement of the pressure over $45°<\theta<135°$.

Other important features of the surface response are apparent in Figs. 6 and 7. The radial velocity at $\theta=0$ is essen-

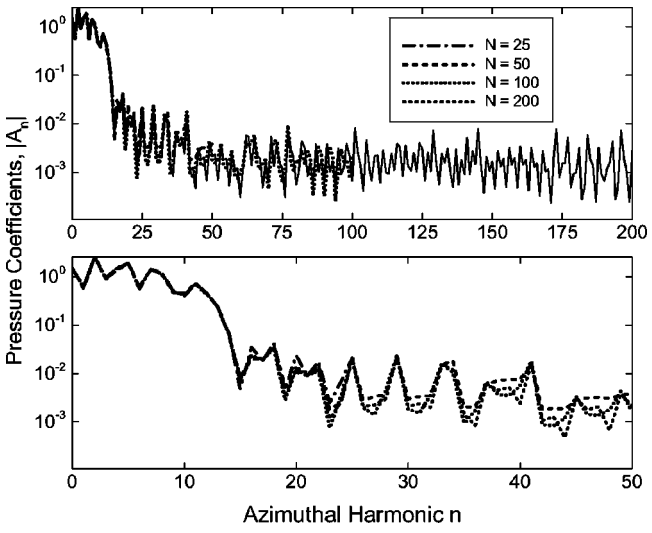


FIG. 4. Dependence of the pressure series coefficients on the azimuthal harmonic number for a partially coated cylinder according to collocation, $z_1=0$, $z_2=100$, Re=300, $ka=10$.

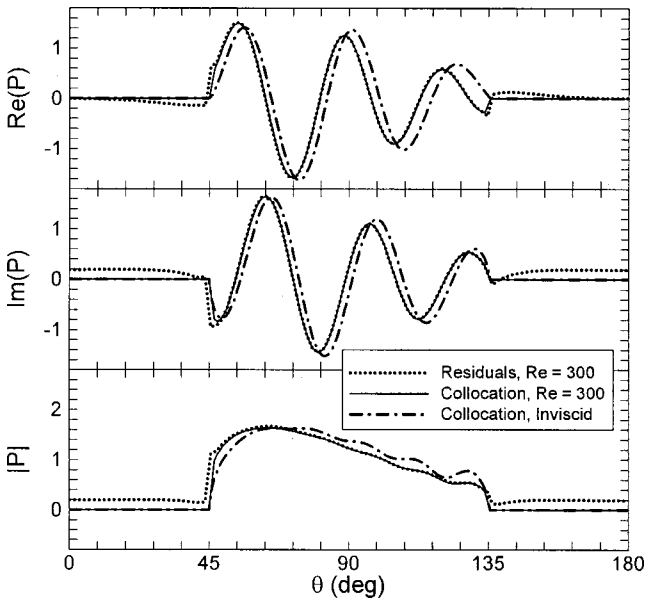


FIG. 6. Surface pressure for a partially coated cylinder obtained by collocation, $z_1=0$, $z_2=100$, $ka=10$.

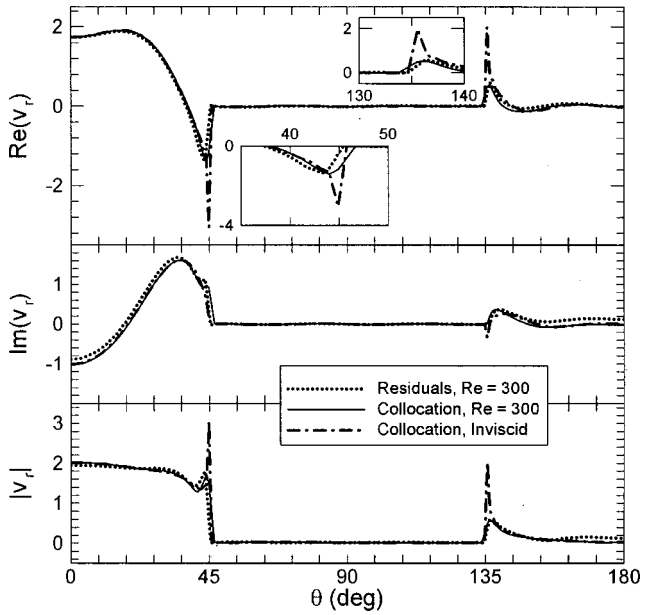


FIG. 7. Radial velocity on the surface of a partially coated cylinder, $z_1=0$, $z_2=100$, $ka=10$.

tially twice the incident value, which is to be expected because at this location the plane wave is normally incident to a pressure–release surface, and the frequency is reasonably high. The radial velocity on the coated region $3\pi/4<\theta<\pi$ is very small because this is the shadowed side of the cylinder. Also, one should note in Fig. 6 that along the uncoated region, which is close to horizontal, the peak values of Re(p) are separated by approximately $\pi/6$, as are the peak values of Im(p). This is approximately the angle subtended by one wavelength of the plane incident wave, $\lambda/a=2\pi/ka$, when $ka=10$.

Figure 8 shows the effect of viscosity on the pressure coefficients obtained by collocation. Viscosity enhances the coefficients for $n<25$, but the differences are relatively small. For $n>25$, there are significant differences in the results for Re=300 and Re=∞, but the coefficients are relatively small in either case. The differences between the two sets of amplitudes lead to an overall enhancement in the farfield pressure, as shown in the directivity plots of Fig. 9. This data is computed by introducing the large-argument asymptotic approximation of the Hankel function, Eq. (22), into the Fourier series for pressure, Eq. (13c). Also, to compensate for scaling the Hankel functions and series coefficients, each value of A_n must be scaled down by the same power of 10, S_n, used to scale down the Hankel functions from which the coefficient equations are formed. The resulting farfield pressure is

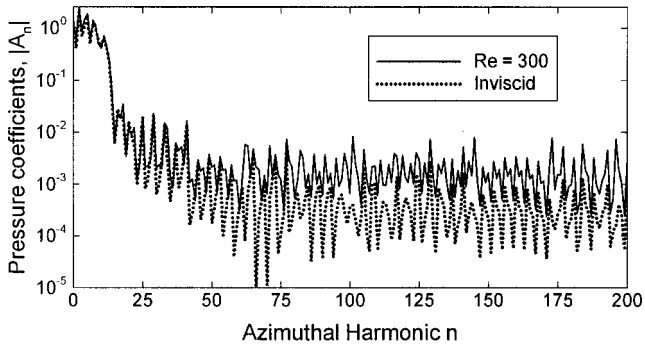


FIG. 8. Effect of viscocity on the pressure series coefficients for a partially coated cylinder according to collocation, $z_1=0$, $z_2=100$, $ka=10$.

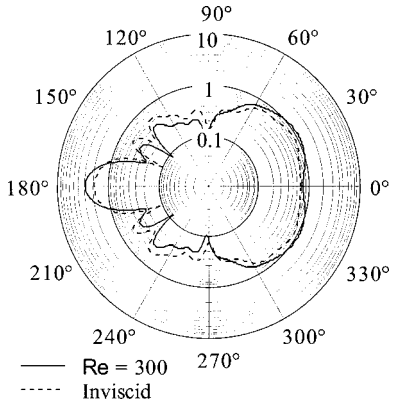


FIG. 9. Directivity of the pressure, $\sqrt{r}|p|$, scattered from a partially coated cylinder according to collocation, $z_1=0$, $z_2=100$, Re=300, $ka=10$.

$$\sqrt{r}p=\left(\frac{2}{\pi\lambda}\right)^{1/2}\exp(i\lambda r)\sum_{n=0}^{N}10^{-S_n}A_n \times\exp\left[-i\left(\frac{2n+1}{4}\right)\pi\right]\cos(n\theta). \tag{23}$$

Overall, viscocity seems to enhance back-scattering and lessen forward-scattering, although the effect is relatively small.

An interesting aspect of the evaluation of the farfield pressure is the fact that there is no discernible difference between the collocation and weighted residual results. The underlying cause for this agreement would be apparent if the pressure coefficients in Figs. 2 and 4 were overlaid. Doing so would reveal that the two sets of data are in close agreement for the first 25 azimuthal harmonics, which are the dominant contributors to radiation.

The surface pressure for other frequencies is depicted in Fig. 10. The oscillations of the real and imaginary parts correspond to the wavelengths of the respective surface waves. The attenuation of the pressure amplitude with increasing θ at $ka=50$ is a manifestation of the shadowing effect associated with geometrical acoustics. Viscosity is seen to lead to substantial attenuation at the higher frequency.

The directivity, $\sqrt{r}|p|$ as a function of θ, at $ka=50$ is depicted in Fig. 11. Viscosity reduces the overall level, especially for $\theta<90°$, which is consistent with the lower pressure amplitude on the surface. However, the pattern for Re =300 shows numerous nulls not seen in the inviscid result. The explanation of this effect, which is unexpected, lies in Fig. 12, in which the magnitude of the pressure coefficients have been scaled downward, as in Eq. (23). It is evident in this figure that viscosity attenuates all harmonics other than those around $n=25$ and $n=50$. The reduced contributions of lower harmonics has the effect of enhancing the directivity's fluctuations.

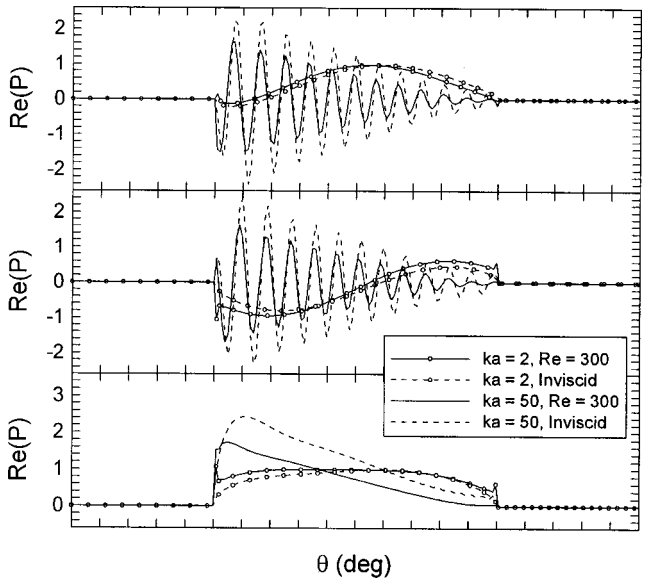


FIG. 10. Effect of frequency on the surface pressure for a partially coated cylinder obtained by collocation, $z_1=0$, $z_2=100$.

VII. CLOSURE

It is clear from the analysis and results that viscosity has a negligible effect on scattering from partially coated bodies, unless the Reynolds number is very small. Its overall effect is to attenuate the surface pressure, and to smooth out the singularity in surface normal velocity that arises in an ideal fluid at impedance discontinuities. From the viewpoint of the acoustic field surrounding the cylinder, viscosity has two effects. The presence of viscosity alters the amplitude coefficients describing the acoustical modes. This alters the pressure field everywhere. In contrast, the contribution of the vortical modes is confined to a thin boundary layer. This is indicated in the analysis by the fact that the Hankel function radial dependence of the vorticity modes, in which the radial dependence of each azimuthal harmonic depends on $H_n(e^{i\pi/4}\eta r)$, where $\eta=(ka\,\mathrm{Re})^{1/2}$. Because Re is large, the magnitude of this term is approximately $\exp[-(ka\,\mathrm{Re}/2)(r-1)]$, where ar is the dimensional radial distance. At radial distances from the surface, $a(r-1)$, that are a few multiples of $2/(k\,\mathrm{Re})$, all that is left are the acoustic modes.

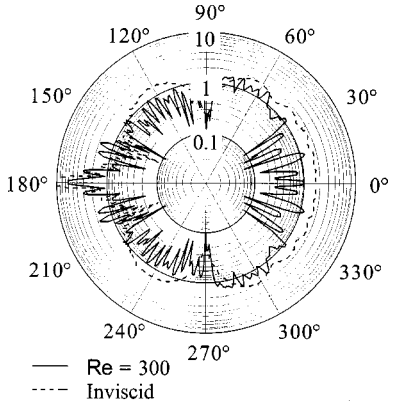


FIG. 11. Directivity of the pressure, $\sqrt{r}|p|$, scattered from a partially coated cylinder according to collocation, $z_1=0$, $z_2=100$, Re=300, $ka=50$.

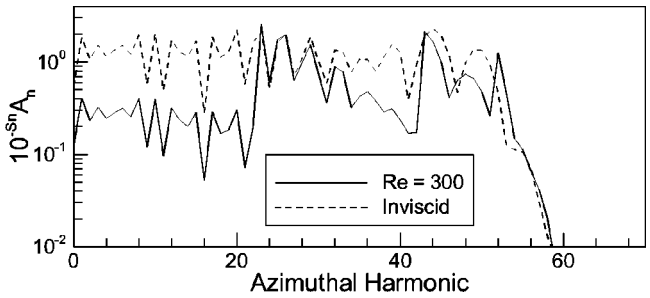


FIG. 12. Scaled pressure coefficient magnitudes, $ka=50$.

A general conclusion that one may draw from this study is that weighted residual methods are not reliable for analyzing the surface response in situations where the boundary conditions are spatially discontinuous. In essence, although the method minimizes the average difference between the approximate and true solutions, it still permits this difference at any specific location to be large either positively or negatively. Some problems of this type, particularly those featuring an infinite half-space, lend themselves to analysis using the Weiner–Hopf transformation. In other situations, boundary and finite element formulations, which attempt to enforce the governing laws over small subregions, are likely to fare better than Ritz expansion procedures. On the other hand, if one is solely interested in processes where the lower harmonics are significant, for example, the farfield pressure in the present situation, then weighted residuals solution might be sufficiently accurate.

ACKNOWLEDGMENTS

The author is grateful for the assistance of James Gregory McDaniel of Boston University in the early stages of the investigation when he was a graduate student at the Georgia Institute of Technology.

[1] A. A. Ferri, J. H. Ginsberg, and P. H. Rogers, "Scattering of plane waves from submerged objects with partially coated surfaces," J. Acoust. Soc. Am. **92**, 1721–1728 (1992).
[2] E. P. Babailov, "Green's function of the Helmholtz equation for a medium with mixed boundary conditions," Sov. Phys. Acoust. **31**, 88–93 (1985).
[3] A. Büyükaksoy and G. Uzgören, "High-frequency scattering from the impedance discontinuity on a cylindrical surface," IEEE Trans. Antennas Propag. **35**, 234–236 (1987).
[4] C. E. Shaw, "The effects of imperfections on the acoustic scattering of a coated shell," M. S. thesis, Georgia Institute of Technology, 1992.
[5] C. E. Shaw and J. H. Ginsberg, "The effect of incomplete coating coverage on acoustic scattering from cylinders," J. Acoust. Soc. Am. **92**, 2337 (1992).
[6] J. M. Cuchieri and D. Feit, "Influence of circumferential partial coating on the acoustic radiation from a fluid-loaded shell," J. Acoust. Soc. Am. **92**, 1721–1728 (2000).
[7] M. Naghieh and S. I. Hayek, "Diffraction of a point source by two impedance covered half-planes," J. Acoust. Soc. Am. **69**, 629–637 (1981).
[8] D. Habault and P. J. T. Filippi, "Ground effect analysis: surface wave and layer potential representations," J. Sound Vib. **79**, 529–550 (1981).
[9] P. H. Dahl and G. V. Frisk, "Diffraction from the juncture of pressure-release and locally reacting half-planes," J. Acoust. Soc. Am. **90**, 1093–1100 (1991).
[10] J. N. Tjøtta, "Reflection of an acoustic wave on a plane surface presenting an impedance discontinuity along a straight line," Acta Acust. **1**, 15–33 (1993).
[11] M. Hasheminejad and T. L. Geers, "Modal impedances for two spheres in a thermoviscous fluid," J. Acoust. Soc. Am. **94**, 2205–2214 (1993).

[12] A. D. Pierce, *Acoustics* (McGraw-Hill, New York, 1981), Chap. 10.
[13] C. F. Du Toit, ''Numerical computations of Bessel functions of the first and second kind for integer orders and complex arguments,'' IEEE Trans. Antennas Propag. **38**, 1341–1349 (1990).
[14] J. P. Mason, ''Cylindrical Bessel functions for a large range of complex arguments,'' Comput. Phys. Commun. **30**, 1–11 (1983).
[15] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972), Chap. 9.

Aberration correction for time-domain ultrasound diffraction tomography

T. Douglas Mast[a)]
Applied Research Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 9 August 2001; revised 26 March 2002; accepted 4 April 2002)

Extensions of a time-domain diffraction tomography method, which reconstructs spatially dependent sound speed variations from far-field time-domain acoustic scattering measurements, are presented and analyzed. The resulting reconstructions are quantitative images with applications including ultrasonic mammography, and can also be considered candidate solutions to the time-domain inverse scattering problem. Here, the linearized time-domain inverse scattering problem is shown to have no general solution for finite signal bandwidth. However, an approximate solution to the linearized problem is constructed using a simple delay-and-sum method analogous to ''gold standard'' ultrasonic beamforming. The form of this solution suggests that the full nonlinear inverse scattering problem can be approximated by applying appropriate angle- and space-dependent time shifts to the time-domain scattering data; this analogy leads to a general approach to aberration correction. Two related methods for aberration correction are presented: one in which delays are computed from estimates of the medium using an efficient straight-ray approximation, and one in which delays are applied directly to a time-dependent linearized reconstruction. Numerical results indicate that these correction methods achieve substantial quality improvements for imaging of large scatterers. The parametric range of applicability for the time-domain diffraction tomography method is increased by about a factor of 2 by aberration correction. [DOI: 10.1121/1.1481063]

PACS numbers: 43.20.Fn, 43.80.Qf, 43.60.Pt [LLT]

I. INTRODUCTION

This paper concerns time-domain diffraction tomography methods for solution of the time-domain inverse scattering problem, in which an unknown inhomogeneous medium is determined from its far-field acoustic scattering. This problem is of interest for medical ultrasonic imaging, since inverse scattering methods such as diffraction tomography can provide quantitative reconstruction of tissue properties including sound speed, density, and absorption.

Most practical inverse scattering methods to date have been based on linearization of the inverse problem using the Born or Rytov approximation.[1,2] These are weak scattering approximations, in which the variation of medium properties is assumed to be a small perturbation from a uniform background. Nonlinear inverse scattering methods,[3,4] which consider contributions of strong and multiple scattering, are much more complex and computationally intensive. However, since large-scale tissue structures cannot be considered weak scatterers at diagnostic ultrasound imaging frequencies,[5,6] linearized inverse scattering methods are of limited use for medical ultrasonic imaging.

A similar problem arises in conventional B-scan and synthetic-aperture imaging,[7,8] which form the basis for current diagnostic ultrasound scanners. Current scanners form synthetic images based on the assumption of a uniform background sound speed, which is essentially the Born approximation. The invalidity of this assumption is associated with image artifacts and focus aberration.[5,9] Considerable effort has been devoted to methods for aberration-corrected imaging, which is analogous to nonlinear inverse scattering. Approaches to aberration correction for pulse-echo imaging have been designed to correct distortion associated with several simplified propagation models, including refraction by homogeneous layers,[10,11] phase aberration close to the transducer aperture,[12–14] and aberration caused by a hypothetical phase screen away from the aperture.[15–17] All of these aberration correction methods require indirect estimation of the medium-induced distortion based on the received scattering data.

A time-domain diffraction tomography method has been introduced recently.[18,19] This method provides tomographic reconstructions of unknown scattering media from scattering data measured on a surface surrounding the region of interest, using the entire available bandwidth of the signals employed. The reconstruction algorithm is derived as a simple delay-and-sum formula similar to synthetic-aperture algorithms employed in conventional clinical scanners.[7,8] However, unlike current clinical scanners, the present method provides quantitative images of the spatially dependent tissue sound speed. These quantitative sound speed maps offer considerable potential for aberration correction, since the medium-induced distortion can be estimated directly from the image data.

The image reconstruction algorithm of Ref. 18 was derived from the frequency-domain exact solution to the linearized inverse scattering problem, i.e., diffraction tomography employing the Born approximation. Inverse scattering approaches based on the Born approximation form adequate

a)Current address: Ethicon Endo-Surgery, 4545 Creek Rd., ML 40, Cincinnati, OH 45242. Electronic mail: dmast@eesus.jnj.com

images only for relatively small, weakly scattering objects,[18,20] so that this approximation has limited utility for large-scale imaging problems such as ultrasonic mammography. In the present paper, an aberration correction approach, which significantly extends the range of validity of the time-domain diffraction tomography method, is introduced. The reconstruction method of Ref. 18 is shown to result in an approximate solution to the time-domain linearized inverse scattering problem; application of aberration correction results in reconstructions that better approximate the solution to the full nonlinear time-domain inverse problem.

Two related methods for aberration correction are presented here. The first, suggested by the synthetic-aperture nature of the reconstruction algorithm, employs a focus correction approach in which delays are computed from estimates of the medium using an efficient straight-ray approximation. The second approach is suggested by examination of the reconstruction itself in the time domain, as in Ref. 21. In this approach, delays are applied directly to a time-dependent linearized reconstruction. Numerical results show that both methods increase the parameter range for which valid images can be obtained and illustrate differences in performance between the two.

II. THEORY

The imaging problem considered here concerns reconstruction of an unknown medium from far-field, time-domain scattering measurements. Solutions of this inverse problem are quantitative images of scattering media such as biological tissue. Below, the linearized inverse scattering problem (e.g., quantitative ultrasonic imaging without aberration correction) is considered and shown to have no general solution. However, approximate solutions to the nonlinear inverse problem result in useful aberration correction methods for quantitative imaging.

A. The linearized time-domain inverse scattering problem

The time-domain inverse scattering problem analyzed below is defined as follows. A quiescent, inhomogeneous, fluid medium is subjected to an incident plane wave pulse propagating in the direction $\boldsymbol{\alpha}$,

$$p_i(\mathbf{r},t)=u(t-\mathbf{r}\cdot\boldsymbol{\alpha}/c_0), \tag{1}$$

where c_0 is a reference or ''background'' sound speed. The medium is assumed to have spatially varying sound speed, constant density, and no absorption, and to be completely characterized by a contrast function $\gamma(\mathbf{r})$, defined as

$$\gamma(\mathbf{r})=\frac{c_0^2}{c(\mathbf{r})^2}-1, \tag{2}$$

where $c(\mathbf{r})$ is the local sound speed at position $\mathbf{r}$. The inverse scattering problem is the determination of the medium contrast $\gamma(\mathbf{r})$ from time-domain measurements of the scattered field $p_s(\boldsymbol{\theta},\boldsymbol{\alpha},t)$ for all measurement directions $\boldsymbol{\theta}$, incident-wave directions $\boldsymbol{\alpha}$, and times t. The implicit neglect of density variations is not severely limiting, since the contrast given by Eq. (2) typically dominates reconstructed images even in the presence of density variations.[18]

A general time-domain solution for the scattered acoustic pressure at a far-field measurement radius R, valid for two-dimensional (2D) or three-dimensional (3D) scattering, is then

$$p_s(\boldsymbol{\theta},\boldsymbol{\alpha},t)=\mathbf{F}^{-1}[\hat{p}_s(\boldsymbol{\theta},\boldsymbol{\alpha},f)]\equiv\int_{-\infty}^{\infty}\hat{p}_s(\boldsymbol{\theta},\boldsymbol{\alpha},f)e^{-i2\pi ft}\,df, \tag{3}$$

where $\hat{p}_s(\boldsymbol{\theta},\boldsymbol{\alpha},f)$ is a single frequency component of the scattered wavefield, given in the far field by

$$\begin{aligned}\hat{p}_s(\boldsymbol{\theta},\boldsymbol{\alpha},f)&=\mathbf{F}[p_s(\boldsymbol{\theta},\boldsymbol{\alpha},t)]d\\&\equiv\int_{-\infty}^{\infty}p_s(\boldsymbol{\theta},\boldsymbol{\alpha},t)e^{i2\pi ft}\,dt\\&=k^2\Gamma(R,f)\int_{V_0}e^{-ik\boldsymbol{\theta}\cdot\mathbf{r}}\gamma(\mathbf{r}_0)\hat{p}_u(\mathbf{r},\alpha,\omega)\,dV_0.\end{aligned} \tag{4}$$

In Eq. (4), k is the wave number ω/c_0 and $\hat{p}_u(\mathbf{r}_0,\boldsymbol{\alpha},\omega)$ is the total frequency-domain acoustic pressure associated with an incident plane wave $\hat{u}(f)e^{ik\boldsymbol{\alpha}\cdot\mathbf{r}_0}$ [i.e., one frequency component of the plane wave pulse $u(t-\boldsymbol{\alpha}\cdot\mathbf{r}/c_0)$]. The integral in Eq. (4) is taken over the entire support of γ in $\mathbb{R}^2$ for 2D scattering or in $\mathbb{R}^3$ for 3D scattering. The term $\Gamma(R,f)$, associated with the far-field forms of the free-space Green's functions for the Helmholtz equation,[22] is

$$\begin{aligned}\Gamma(R,f)&=-\sqrt{\frac{i}{8\pi kR}}\quad\text{for } 2D \text{ scattering,}\\\Gamma(R,f)&=\frac{1}{4\pi r}\quad\text{for } 3D \text{ scattering.}\end{aligned} \tag{5}$$

The time-domain inverse scattering problem is given by the Fourier inverse of Eq. (4):

$$p_s(\boldsymbol{\theta},\boldsymbol{\alpha},t)=\int_V\mathbf{L}\left[p_u\left(\mathbf{r},\boldsymbol{\alpha},t-\frac{R}{c_0}+\frac{\boldsymbol{\theta}\cdot\mathbf{r}}{c_0}\right)\right]\gamma(\mathbf{r})\,dV, \tag{6}$$

where $p_u(\mathbf{r},\boldsymbol{\alpha},t)$ is the total time-domain acoustic pressure associated with the incident plane wave $u(t-\mathbf{r}\cdot\boldsymbol{\alpha}/c_0)$ and the linear operator $\mathbf{L}$ is defined as

$$\begin{aligned}&\mathbf{L}[p(\mathbf{r},\boldsymbol{\alpha},t)]\\&\quad=\frac{1}{c_0^2}\mathbf{F}^{-1}\left[\sqrt{\frac{i}{8\pi kR}}\mathbf{F}[\ddot{p}(\mathbf{r},\boldsymbol{\alpha},t)]\right]\quad\text{for } 2\text{D scattering,}\\&\mathbf{L}[p(\mathbf{r},\boldsymbol{\alpha},t)]=-\frac{1}{4\pi c_0^2R}\ddot{p}(\mathbf{r},\boldsymbol{\alpha},t)\quad\text{for } 3\text{D scattering.}\end{aligned} \tag{7}$$

Equation (6) defines a nonlinear inverse problem for the contrast $\gamma(\mathbf{r})$; the nonlinearity is associated with the dependence of $p(\mathbf{r},\boldsymbol{\alpha},t)$ on $\gamma(\mathbf{r})$.

The nonlinear time-domain inverse scattering problem defined by Eq. (6) can be linearized by invoking the Born approximation, in which the total acoustic pressure is approximated by the incident wave. The resulting linearized equation is

$$p_s(\boldsymbol{\theta},\boldsymbol{\alpha},t)=\int_V \mathbf{L}[u(t-\tau(\boldsymbol{\theta},\boldsymbol{\alpha},\mathbf{r}))]\,\gamma_L(\mathbf{r})\,dV, \tag{8}$$

where the true potential $\gamma(\mathbf{r})$ has been replaced by $\gamma_L(\mathbf{r})$, a hypothetical solution to the linearized inverse problem, and the propagation delay term $\tau(\boldsymbol{\theta},\boldsymbol{\alpha},\mathbf{r})$ is defined

$$\tau(\boldsymbol{\theta},\boldsymbol{\alpha},\mathbf{r})\equiv\frac{R}{c_0}-\frac{(\boldsymbol{\theta}-\boldsymbol{\alpha})\cdot\mathbf{r}}{c_0}. \tag{9}$$

The delay specified by Eq. (9) is precisely that required to refocus scattered waves through a homogeneous ($c=c_0$) medium onto each image point.

In the asymptotic weak scattering limit, the linearized inverse scattering problem (8) is equivalent to the original nonlinear problem (6), so that an exact solution for any waveform $u(t)$ is given by $\gamma_L(\mathbf{r})\rightarrow\gamma(\mathbf{r})$ as $\gamma(\mathbf{r})\rightarrow 0$. However, unlike the frequency-domain linearized inverse scattering problem, the inverse problem of Eq. (8) has no general solution for nonzero $\gamma(\mathbf{r})$. To prove this, one may examine the Fourier transform of Eq. (8), which is simply the linearization of Eq. (4):

$$\hat{p}_s(\boldsymbol{\theta},\boldsymbol{\alpha},f)=\Gamma(R,f)\hat{u}(f)\int_V e^{-ik(\boldsymbol{\theta}-\boldsymbol{\alpha})\cdot\mathbf{r}}\gamma_L(\mathbf{r})\,dV, \tag{10}$$

where k is the wave number $2\pi f/c_0$. Thus, any general time-independent solution of Eq. (8) must also be a frequency-independent solution to the linearized frequency-domain inverse scattering problem (10).

For $\boldsymbol{\theta}=\boldsymbol{\alpha}$ (the forward scattering case), Eq. (10) leads to the condition

$$\frac{\hat{p}_s(\boldsymbol{\theta},\boldsymbol{\theta},f)e^{-ikR}}{k^2\hat{u}(f)}=\frac{1}{4\pi R}\int_V \gamma_{\mathbf{L}}(\mathbf{r})dV=\text{const}\ (\forall f) \tag{11}$$

for existence of a general solution to Eq. (8). This requirement is easily seen by counterexample to be impossible. For example, Eq. (11) requires that, for all frequencies f, the magnitude of the forward scattered pressure should (for a unit-amplitude incident wave) be proportional to f^2. A counterexample is given by any high-contrast scatterer (e.g., $\gamma\sim 1$), for which this f^2 dependence occurs only at very low frequencies, such that the scatterer's dimensions are much smaller than the wavelength c_0/f.[23] Thus, although the nonlinear time-domain inverse scattering problem has an exact solution [equal to the true contrast $\gamma(\mathbf{r})$], the corresponding linearized problem has no general solution for arbitrary signal bandwidth except in the limiting case $\gamma\rightarrow 0$.

B. Approximate linearized solutions by Fourier synthesis

Although no solution $\gamma_L(\mathbf{r})$ to the quantitative imaging problem of Eq. (8) exists in general, one can still obtain approximate solutions by applying Fourier synthesis to the well-known exact solution of the frequency-domain linearized inverse scattering problem. For any frequency component of $p_s(\boldsymbol{\theta},\boldsymbol{\alpha},t)$, the frequency-domain linearized inverse problem (10) has an exact, frequency-dependent solution given by the frequency-domain filtered backpropagation formula.[2,24]

$$\gamma_B(\mathbf{r},f)=\frac{\hat{\mu}(f)e^{-ikR}}{\hat{u}(f)}\int\int \Phi(\boldsymbol{\theta},\boldsymbol{\alpha})\hat{p}_s(\boldsymbol{\theta},\boldsymbol{\alpha},f)\times e^{ik(\boldsymbol{\theta}-\boldsymbol{\alpha})\cdot\mathbf{r}}\,dS_\alpha\,dS_\theta, \tag{12}$$

where

$$\hat{\mu}(f)=\sqrt{\frac{kR}{8i\pi^3}},\quad \Phi(\boldsymbol{\theta},\boldsymbol{\alpha})=|\sin(\theta-\alpha)|\ \text{ in 2D},$$
$$\hat{\mu}(f)=\frac{kR}{4\pi^3},\quad \Phi(\boldsymbol{\theta},\boldsymbol{\alpha})=|\boldsymbol{\theta}-\boldsymbol{\alpha}|\ \text{ in 3D}. \tag{13}$$

Each surface integral in Eq. (12) is performed over the entire measurement circle for the 2D case and over the entire measurement sphere for the 3D case.

Fourier inversion of Eq. (10) into the time domain can be performed using the convolution theorem.[25] The result, with the hypothetical linearized solution $\gamma_L(\mathbf{r})$ replaced by the Born reconstruction $\hat{\gamma}_B(\mathbf{r},f)$, is

$$p_s(\boldsymbol{\theta},\boldsymbol{\alpha},t)=-\frac{1}{4\pi c_0^2 R}\int_V \ddot{u}(t-\tau(\boldsymbol{\theta},\boldsymbol{\alpha},\mathbf{r}))\otimes\gamma_B(\mathbf{r},t)\,dV, \tag{14}$$

where $\gamma_B(\mathbf{r},t)$ is the inverse Fourier transform of the frequency-domain solution $\hat{\gamma}_B(\mathbf{r},f)$. The time-domain reconstruction $\gamma_B(\mathbf{r},t)$ is an exact solution of the integral equation (14), which is similar but not equivalent to the linearized time-domain inverse scattering problem of Eq. (8). Because $\hat{\gamma}_B(\mathbf{r},f)$ is conjugate symmetric, the time-domain potential $\gamma_B(\mathbf{r},t)$ is purely real.[21]

Comparison of Eqs. (8) and (14) shows that, in the weak scattering limit,

$$\gamma_B(\mathbf{r},t)\rightarrow\gamma(\mathbf{r})\,\delta(t)+\psi(\mathbf{r},t), \tag{15}$$

where $\psi(\mathbf{r},t)$ is a "nonradiating source"[26] that satisfies the constraint

$$\int_V \ddot{u}(t-\tau(\boldsymbol{\theta},\boldsymbol{\alpha},\mathbf{r}))\otimes\psi(\mathbf{r},t)\,dV=0. \tag{16}$$

The presence of the nonradiating source term $\psi(\mathbf{r},t)$ is consistent with the nonuniqueness of solutions to Eq. (14).[27] For example, additional solutions to Eq. (14) include the class of functions $\gamma_B(\mathbf{r},t)+\phi(\mathbf{r})$, where $\phi(\mathbf{r})$ is the inverse Fourier transform of any function $\hat{\phi}(\mathbf{k})$ that is zero inside the Ewald sphere,[1] defined for the upper frequency limit of the incident pulse as $\mathbf{k}\leqslant 4\pi f_h/c_0$, where f_h is the upper limit of the pulse frequency content.

A straightforward approach to estimate $\gamma_B(\mathbf{r},t)$ [and thus $\gamma(\mathbf{r})$] is to perform inverse Fourier transformation on the frequency-domain Born inversion $\hat{\gamma}_B(\mathbf{r},f)$. A natural estimate of the medium contrast is a reconstruction employing information from multiple frequencies contained in the incident pulse, e.g.,

$$\gamma_v(\mathbf{r},t)=\int_{-\infty}^{\infty}\hat{\gamma}_B(\mathbf{r},f)\hat{v}(f)e^{-2\pi ift}\,df\Big/\int_{-\infty}^{\infty}\hat{v}(f)\,df \tag{17}$$

$$=\gamma_B(\mathbf{r},t)\otimes v(t)/v(0) \tag{18}$$

$$\approx\gamma(\mathbf{r})v(t)/v(0), \tag{19}$$

where the final expression results from Eq. (15). The frequency weight $\hat{v}(f)$ must be integrable and have no support outside the support of $\hat{u}(f)$, but is otherwise arbitrary. The time dependence of the reconstructed contrast can be removed from Eq. (19) by setting $t=0$ (called the "imaging condition" in Ref. 28).

If the incident waveform is sinusoidal, so that, for instance, $\hat{u}(f)=\delta(f-f_0)+\delta(f+f_0)$, the reconstructed potential $\gamma_v(\mathbf{r},0)$ is equal to the real part of the frequency-domain solution $\hat{\gamma}_B(\mathbf{r},f_0)$. Thus, $\gamma_v(\mathbf{r},0)$ is an exact solution of the linearized inverse problem in the single-frequency limit. However, as proven above, no time-independent reconstruction can solve the general linearized time-domain inverse scattering problem, so that $\gamma_v(\mathbf{r},0)$ is only an *approximate* solution for any nonzero-bandwidth incident waveform $u(t)$.

The Fourier inversion of Eq. (17) can be performed either numerically or analytically. Numerical inversion, using frequency-domain reconstructions at a number of discrete frequencies within the bandwidth of the incident pulse, was the approach employed by Lin, Nachman, and Waag.[21] (However, the frequency-domain inversions of Ref. 21 were performed using eigenfunctions of the far-field scattering operator[29] instead of filtered backpropagation.) Alternatively, particular choices of the weight $\hat{v}(f)$ allow analytic inversion of the frequency-domain reconstruction $\gamma_B(\mathbf{r},f)$ into the time domain, resulting in a simple delay-and-sum formula. For the weight $\hat{v}(f)=\hat{u}(f)/\hat{\mu}(f)H(f)$, where $H(f)$ is the Heaviside step function, the resulting formula is

$$\gamma_v(\mathbf{r},t)=\mathrm{Re}\left[\frac{1}{N}\int\int\Phi(\boldsymbol{\theta},\boldsymbol{\alpha})(p_s(\boldsymbol{\theta},\boldsymbol{\alpha},\tau)+i\mathbf{H}^{-1}[p_s(\boldsymbol{\theta},\boldsymbol{\alpha},\tau)])\,dS_\alpha\,dS_\theta\right], \tag{20}$$

where

$$N=2\int_0^{\infty}\frac{\hat{\mu}(f)}{\hat{\mu}(f)}df, \tag{21}$$

τ is given by Eq. (9), and $\mathbf{H}^{-1}$ is the inverse Hilbert transform operator (quadrature filter), which results from limiting frequency integration to the interval $(0,\infty)$.[18]

The reconstruction formula of Eq. (20) is identical to that derived in Ref. 18 and similar to that derived in Ref. 30. In view of the present derivation, these previous methods are understood to provide approximate solutions to the linearized time-domain inverse scattering problem (8).

C. Aberration-corrected solutions

The form of the approximate linearized solution derived above suggests possible approaches to improvement of images beyond the limits of the Born approximation.

First, one may observe that the reconstruction formula of Eq. (20) synthetically focuses the time-domain scattered field back onto each point in the medium.[18] This observation leads to the idea of aberration correction by iterative refinement of the focus quality. Since the reconstruction provides an estimate of the medium itself, this refinement is fairly straightforward. One simple implementation employs an assumption that background inhomogeneities result only in cumulative delays (or advances) of the incident and scattered wavefronts, so that the total delay for an angle ϕ and a point position $\mathbf{r}$ is given by

$$\delta\tau(\boldsymbol{\phi},\mathbf{r})=\int_{\xi}c(\boldsymbol{\xi})^{-1}d\xi-\frac{R}{c_0}, \tag{22}$$

where the integral is performed along the line that joins the spatial points $\mathbf{r}$ and $R\boldsymbol{\phi}$, Aberration-corrected reconstructions can then be performed using Eq. (20) with τ replaced by the corrected delay term

$$\tau\rightarrow R/c_0+\frac{(\boldsymbol{\alpha}-\boldsymbol{\theta})\cdot\mathbf{r}}{c_0}+\delta r(\boldsymbol{\alpha},\mathbf{r})+\delta\tau(\boldsymbol{\theta},\mathbf{r}) \tag{23}$$

and by then computing $\gamma_v(\mathbf{r},0)$ using Eq. (20).

An alternative approach to aberration correction is motivated by the observation, made in Ref. 21, that temporal delays from wave propagation in the inhomogeneous medium result in corresponding delays to the time-domain reconstruction of Eq. (17). That is, the reconstructed waveforms $\gamma_v(\mathbf{r},t)$ may be delayed or advanced relative to the waveform $v(t)$. In Ref. 21, correction for this temporal aberration was implemented by adaptive demodulation of $\gamma_v(\mathbf{r},t)$ from the weighting waveform $v(t)$. Here, envelope detection is applied to $\gamma_v(\mathbf{r},t)$ and the time of maximum envelope amplitude $t_{\max}$ is found for each point $\mathbf{r}$, resulting in the aberration-corrected reconstruction

$$\gamma(\mathbf{r})\approx\gamma_v(\mathbf{r},t_{\max}(\mathbf{r})). \tag{24}$$

Envelope detection can also be applied to iterative reconstructions obtained using the focus correction given by Eq. (22).

III. COMPUTATIONAL METHODS

The present aberration correction methods have been tested using simulated scattering data for a number of two-dimensional test objects. The computational configuration was chosen to mimic the characteristics of an available 2048-element ring transducer.[31] The time-domain waveform employed for all the computations reported here was

$$u(t)=\cos(\omega_0 t)e^{-t^2/(2\sigma^2)}, \tag{25}$$

where $\omega_0=2\pi f_0$ for a center frequency of f_0, taken here to be 2.5 MHz, and σ is the temporal Gaussian parameter. The value of σ chosen here was 0.25, which corresponds to a -6-dB bandwidth of 1.5 MHz.

For 2D cylindrical inhomogeneities, the frequency-domain scattered field $\hat{p}_s(\boldsymbol{\theta},\boldsymbol{\alpha},\omega)$ was computed using an exact series solution[32] for each frequency component of interest. In implementation of the series solution, summations were truncated when the magnitude of a single coefficient

dropped below 10^{-12} times the sum of all coefficients. These single-frequency solutions, which correspond to Fourier coefficients of the time-domain scattered field, were weighted and inverted by discrete Fourier transform to obtain the exact time-domain scattered field associated with the incident pulse of Eq. (25). Scattering from cylinders of radius 4.0 mm and contrasts ranging from $\gamma=0.001$ to $\gamma=0.14$ was computed on a measurement circle of radius 176 mm for 384 incident-wave directions and 96 measurement directions. The sampling rate employed was 9.14 MHz.

Solutions were also obtained for a large-scale breast model using a time-domain *k*-space method.[33] The breast model was obtained by image processing a coronal cross section of three-dimensional photographic data from the Visible Human Female data set with a pixel size of 0.333 mm. Hue, saturation, and value were mapped to sound speed and density using empirically determined relations. Sound speed and density were assumed to be linearly proportional; this assumption is realistic for mammalian soft tissues.[34,35]

Sound speed and density maps were smoothed using a Gaussian filter to reduce artifacts associated with the slicing process. (The tissue map employed is shown in Fig. 3.) This tissue model was scaled down by a factor of 0.6 from the original data set and mapped onto a grid of 512×512 points with a spatial step of 0.111 mm. A time step of 0.0546 μs, corresponding to a Courant–Friedrichs–Lewy number of 0.75, was employed. Based on the scaling of the tissue model, the scattered field obtained is equivalent to that of the full-scale breast model (largest dimension 75 mm) for a center frequency of 0.5 MHz.

Scattered acoustic pressure signals were recorded at a sampling rate of 9.15 MHz for 128 incident-wave directions. A circle of 512 simulated point receivers, which had a radius of 9.0 mm in these computations (equivalent to a radius of 45 mm for a 0.5 MHz center frequency), completely contained the scaled-down breast model. Far-field waveforms were computed by Fourier transforming the time-domain waveforms on the near-field measurement circle, transforming these to far-field waveforms for each frequency using a numerically exact transformation method,[21,36] and performing inverse Fourier transformation to yield time-domain far-field waveforms at a measurement circle of radius 234 cm (or 1170 mm if scaled to a 0.5 MHz center frequency). All forward and inverse temporal Fourier transforms, as well as angular transforms occurring in the near-field–far-field transformation, were performed by fast Fourier transforms (FFTs).[37]

The time-domain imaging method was directly implemented using Eq. (20), evaluated using straightforward numerical integration over all incident-wave and measurement directions employed. In one implementation, similar to that from Ref. 18, images were evaluated only for the time $t=0$. In this case, before evaluation of the argument τ for each signal, the time-domain waveforms were resampled at a sampling rate of 16 times the original rate. This resampling was performed using FFT-based Fourier interpolation. The inverse Hilbert transform was implicitly performed using the same FFT operation. Values of the pressure signals at the time τ were then determined using linear interpolation between samples of the oversampled waveforms. The integrals of Eq. (20) were implemented using discrete summation over all transmission and measurement directions employed.

In the implementation of reconstructions for multiple times, storage and computation time requirements necessitated modification of the algorithm implementation. For multiple-time reconstructions, a reconstruction of $\gamma_v(\mathbf{r},t)$ at the sampling rate of the scattering data was first obtained by direct integration. Delays of the time-domain scattered waveforms were implemented using cubic spline interpolation.[38] Reconstructions were performed for an interval of length 2.4 μs, multiplied by a window with cosine tapers of length 0.6 μs at each end, and upsampled by a factor of 8 using Fourier interpolation. Inverse Hilbert transformation of $\gamma_v(\mathbf{r},t)$ was performed by the same FFT operation used to implement the Fourier interpolation. Finally, the temporal position of the envelope peak was found from the zero crossing of the envelope derivative,

$$\frac{\partial\left|\gamma_v(\mathbf{r},t_{\text{peak}})+iH^{-1}[\gamma_v(\mathbf{r},t_{\text{peak}})]\right|}{\partial t}=0. \tag{26}$$

The derivative in Eq. (26) was evaluated using a second-order-accurate center-difference scheme.

Focus correction was implemented using a straight-ray approximation, which is based on the assumption that background inhomogeneities result only in cumulative delays (or advances) of the incident and scattered wavefronts. In this approximation, the total delay for an image position $\mathbf{r}$ and a direction $\boldsymbol{\phi}$ is given by Eq. (22) and aberration-corrected reconstructions are performed using Eq. (20) with τ replaced by the corrected delay term of Eq. (23). The path integrals of Eq. (23) were performed using an algorithm based on the digital differential analyzer (DDA) image processing method.[39] This method very efficiently finds the nearest neighbors to a line of specified starting position and slope; thus, the integrals can be evaluated by simple summation without any need for interpolation. To account for variable step size along the integration path, this summation is normalized by multiplication with L/N, where L is the length of the specified line and N is the number of points employed in the summation. Since the reconstruction process acts in part as a low-pass filter, the integral performed using nearest neighbors to the line of interest is sufficiently accurate.

Iterative focus correction was performed by first constructing an uncorrected image, either for $t=0$ or $t=t_{\text{peak}}$. The reconstructed sound speed was then employed to evaluate the delay corrections of Eq. (23) using the DDA implementation of the integrals from Eq. (22). To avoid spurious modification of image points outside the support of the scatterer, the delay term of Eq. (23) was multiplied by the factor

$$A=\begin{cases}1, & |\gamma_v(\mathbf{r})|\geq\gamma_{\max}/2,\\ (1-\cos[2\pi|\gamma_v(\mathbf{r})|/\gamma_{\max}])/2, & |\gamma_v(\mathbf{r})|<\gamma_{\max}/2,\end{cases} \tag{27}$$

where $\gamma_{\max}$ is the maximum value of $|\gamma_v(\mathbf{r})|$ for the previous reconstruction and the temporal criterion ($t=0$ or $t=t_{\text{peak}}$) employed.

Iteration proceeded as follows. A new reconstruction was compared to the previous reconstruction; if the relative

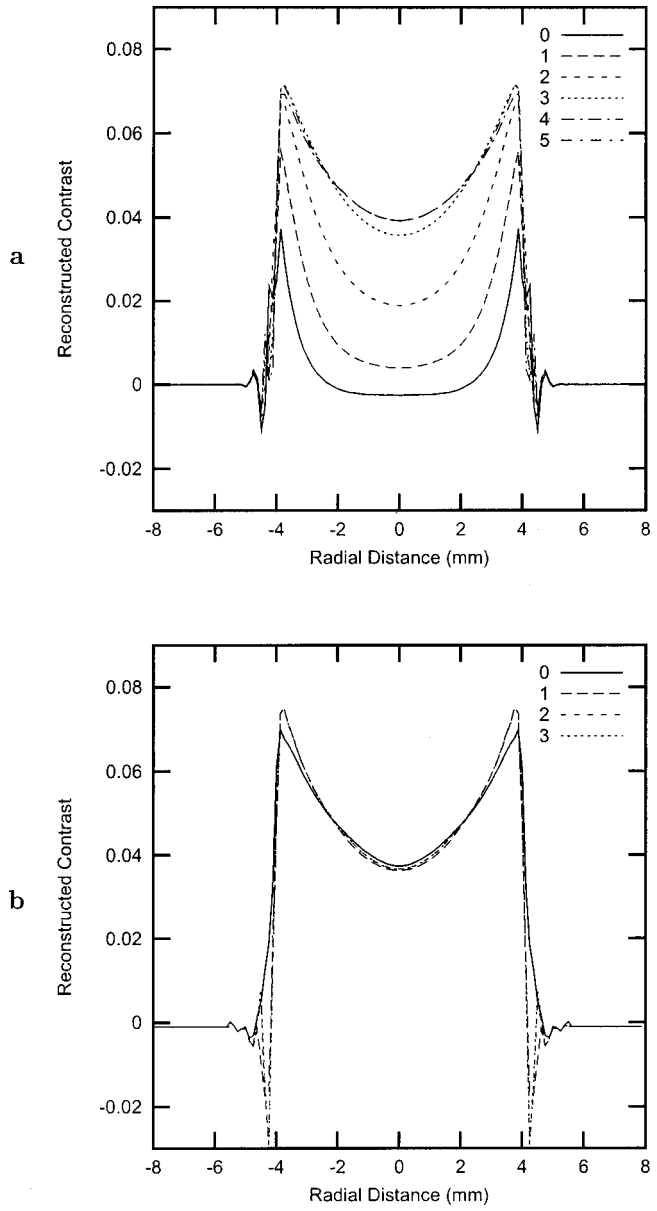


FIG. 1. Cross sections of time-domain reconstructions with adaptive focus correction for both imaging criteria. Reconstructions are of a homogeneous cylinder with a radius of 4 mm ($ka=41.2$) and a contrast $\gamma=0.08$. In each case, the "0" curve refers to an uncorrected reconstruction, while curves labeled "1" and higher correspond to subsequent iterations of focus correction. (a) $t=0$. (b) $t=t_{\text{peak}}$.

rms error between the two was greater than 5%, further iterations were carried out up to a prescribed maximum number of iterations, taken here to be 20. The criterion of 5% was chosen because image quality was not substantially enhanced by use of lower error thresholds. Due to the efficiency of the delay computation, each iteration required about the same computation time as the original reconstruction.

IV. NUMERICAL RESULTS

The performance of aberration-corrected time-domain diffraction tomography imaging, using the two approaches introduced above, is illustrated by the numerical examples presented in this section.

Figure 1 shows reconstructions of a homogeneous cylinder with a radius of 4 mm and a contrast $\gamma=0.08$. For the center frequency of 2.5 MHz, this corresponds to a nondimensional radius $ka=41.2$. Panel (a) shows cross sections of reconstructions obtained using the $t=0$ criterion. The "0" curve refers to an uncorrected (Born approximation) reconstruction, while curves labeled "1" and higher correspond to subsequent iterations of focus correction performed using the delay correction of Eq. (23) as described in Sec. III. Panel (b) shows corresponding cross sections obtained using the $t=t_{\text{peak}}$ criterion. One may observe that iterative focus correction greatly improves reconstructions for the $t=0$ criterion. The initial (Born) reconstruction shows mainly the edges of the cylinder; further iterations improve the accuracy within the cylinder interior. This process somewhat resembles the inverse scattering method of layer stripping,[40,41] in which an unknown medium is iteratively reconstructed with each iteration probing further into the medium interior.

In contrast, iterative focus correction provides little, if any, improvement to the reconstructions obtained using the $t=t_{\text{peak}}$ criterion [Fig. 1(b)]. In this case, the initial reconstruction captures the cylinder interior very well. Further iterations slightly increase the reconstructed contrast near the edges, but also introduce artifacts not present in the initial reconstruction. After convergence, the reconstructed value is more accurate than the $t=0$ image for the cylinder edges but less accurate for the interior.

For the reconstructions shown in Fig. 1, images of size 128×128 pixels were computed from time-domain scattering data for 96 incident-wave directions and 384 measurement directions. The computation time required on a 650-Mhz Athlon processor was about 6 CPU min per iteration for the $t=0$ image criterion (about 38 min total for the six iterations performed) and about 45 CPU min per iteration for the $t=t_{\text{peak}}$ criterion.

The relative performance of iterative focus correction using the two image criteria is illustrated in Fig. 2. Here, reconstructions were based on exact scattering data for a 4-mm cylinder with contrast $0.01\leqslant\gamma\leqslant0.12$. Since previous studies have shown that the accuracy of diffraction tomography reconstructions is roughly a function of the nondimensional parameter $ka\cdot\gamma$,[18,20] the relative error is plotted as a function of this nondimensional parameter. The Born approximation is considered to provide useful images for cylinders up to $ka\cdot\gamma\sim2$;[18,20] by this standard, the iterative focus correction implemented here increases the upper limit of validity for $t=0$ images to $ka\cdot\gamma\sim4$. As in Fig. 1, iterative focus correction is seen to provide little improvement in accuracy for images obtained using the $t=t_{\text{peak}}$ criterion. The quantitative accuracy of reconstructions is slightly increased by iteration for large values of the parameter $ka\cdot\gamma$, but can be slightly diminished for smaller values. Also notable is that iteration using the $t=0$ criterion fails completely above $ka\cdot\gamma\sim4$, while the $t=t_{\text{peak}}$ criterion reaches a comparable error level around $ka\cdot\gamma\sim4$ and then increases gradually in error with increasing scatterer contrast.

Quantitative images of a large-scale 2D breast model, used to generate simulated scattering data in the manner described in Sec. IV, are shown in Figs. 3 and 4. Panel (a) of Fig. 3 shows the 2D model used to generate the synthetic

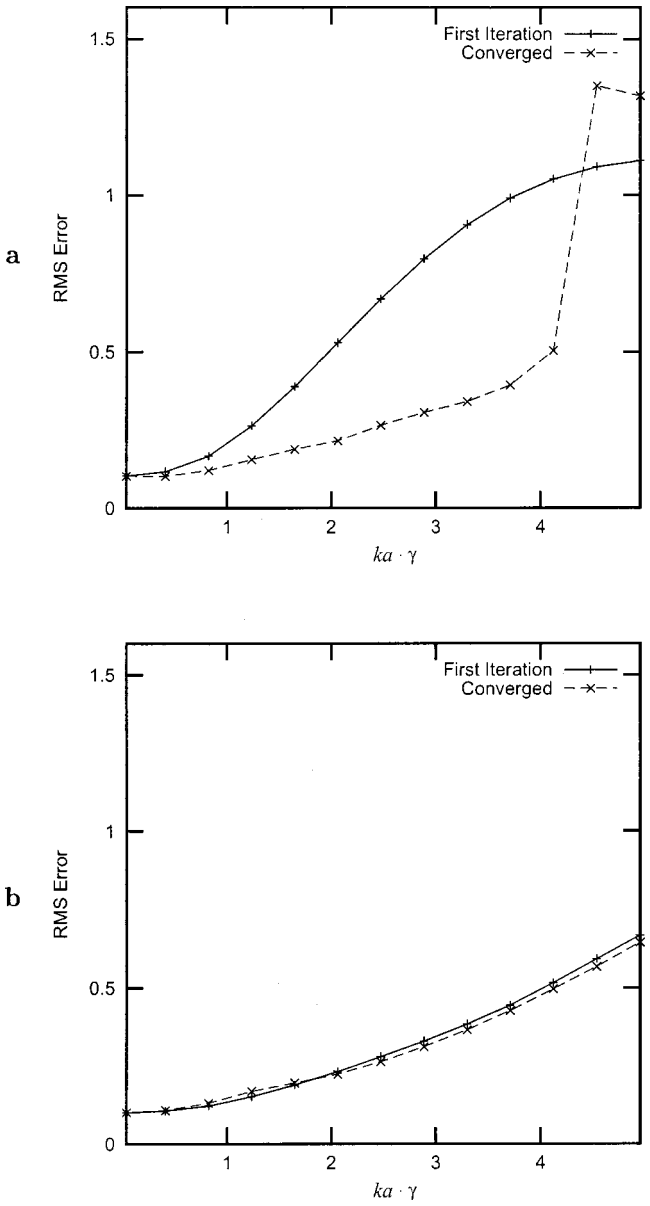


FIG. 2. The rms error for reconstructions of a 4.0-mm-radius cylinder with both imaging criteria, with adaptive focus correction (solid lines) and without (dashed lines). (a) $t=0$. (b) $t=t_{\text{peak}}$.

data. For this model, the parameter $ka \cdot \gamma$ is about 9.3 if estimated using the sound speed of fat, the center frequency of 0.5 MHz, and the largest half-dimension of 37.5 mm. However, a more conservative estimate employing the average contrast γ within the scatterer yields $ka \cdot \gamma \sim 2.8$, which meets the accuracy criterion $ka \cdot \gamma < 4$ determined from the cylinder simulations. Panel (b) of Fig. 3 shows the image reconstructed using the $t=t_{\text{peak}}$ criterion without any focus correction. In this case, the reconstructed image appears to be artifactually sharpened compared to the original model. Although there is a close correspondence between most features of the model and the reconstruction, some differences exist. For example, the reconstructed skin thickness is significantly smaller than that of the actual model in several locations.

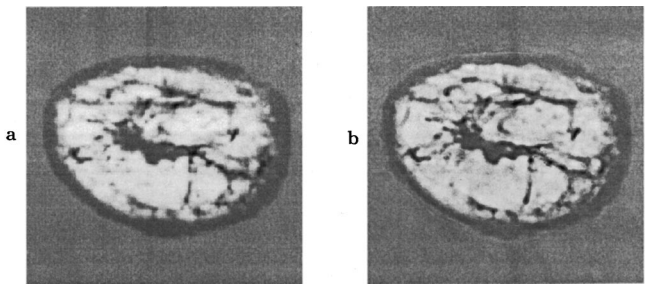


FIG. 3. Reconstruction of a large-scale two-dimensional breast model from simulated scattering data. (a) Model. (b) Initial time-domain reconstruction using $t=t_{\text{peak}}$ criterion.

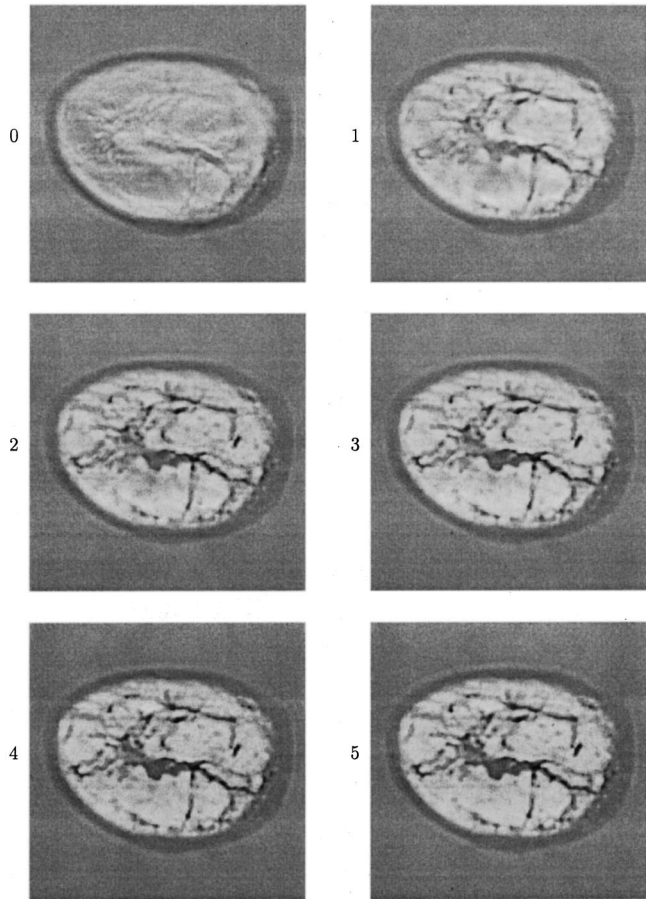


FIG. 4. Images of the large-scale breast model obtained using the $t=0$ criterion with adaptive focusing. Panel 0 shows the initial (linear) reconstruction and panels 1–5 show the subsequent iterations up to convergence.

Reconstructions of the 2D breast model, obtained using the $t=0$ criterion and iterative focus correction, are shown in Fig. 4. In this case, the initial (Born) reconstruction renders the skin layer fairly well, but the interior of the breast model is reconstructed poorly. Subsequent iterations improve the rendering of the connective and glandular tissue structure within the breast. Both focus quality and quantitative accuracy of the reconstructions improve with iteration. The converged reconstruction (iteration 5) resembles a low-pass filtered version of the original model [Fig. 3(a)] except for a small area of spuriously high reconstructed contrast within the interior glandular tissue.

Both reconstruction criteria successfully image the sound speed variation of the 2D breast model, even though the model also included realistic density variations. This re-

sult is expected, since diffraction tomography images of sound speed are not greatly degraded by any density variations that are small and fairly smooth.[18] These criteria are met by the breast model employed here, in which the density variations were of comparable magnitude to the small (maximum about 6%) sound speed variations.

For the large-scale 2D breast model, computation times required for 256×256 pixel images, 128 incident-wave directions, and 512 measurement directions were about 1.3 CPU h per iteration for the $t=0$ image criterion (8.0 h for the six iterations up to convergence) and about 4.6 CPU h for the initial reconstruction using the $t=t_{\text{peak}}$ criterion.

V. DISCUSSION

The two abberation correction methods considered here may be compared as follows. Both methods have the effect of improving the alignment of the time-domain reconstruction $\gamma_v(\mathbf{r},t)$. In the case of $t=0$ images with adaptive focus correction, the time-domain reconstruction is implicitly aligned by compensation for propagation delay within the inhomogeneous medium. The $t=t_{\text{peak}}$ criterion can be thought of as an explicit alignment of the time-domain reconstruction.

Previous qualitative studies of the validity of the Born approximation[18,20] have established a threshold for valid Born reconstructions at $ka\cdot\gamma\sim2$, which corresponds to a normalized rms error of about 0.5 (Fig. 2). Given this somewhat arbitrary threshold for the maximum allowable error, both aberration correction methods employed here have a similar range of validity, up to about $ka\cdot\gamma\sim4$. Thus, either approach extends the parametric range of validity for time-domain diffraction tomography by about a factor of 2.

Each image criterion also introduces characteristic artifacts. The $t=0$ criterion with adaptive focusing acts in part as a low-pass filter to reconstructions, consistent with the well-known low-pass filtering effect of conventional diffraction tomography.[1] The $t=t_{\text{peak}}$ criterion introduces edge artifacts that have the qualitative effect of artifactually sharpening images. More robust methods of delay estimation, such as cross-correlation between the time-domain reconstruction $\gamma_v(\mathbf{r},t)$ and the modulating waveform $v(t)$,[21] may provide better reconstruction quality than the $t=t_{\text{peak}}$ criterion, particularly for scattering data corrupted by noise or measurement imprecision.

The $t=0$ image criterion can provide faster reconstructions, since the reconstructed contrast $\gamma_v(\mathbf{r},t)$ needs only to be evaluated for one time. However, for large or high-contrast scatterers, iterative aberration correction is necessary to obtain high-quality reconstructions. The $t=t_{\text{peak}}$ criterion requires longer computation time for each reconstruction; however, because this criterion implicitly incorporates a form of aberration correction, subsequent iterations provide little additional benefit. As a result, computation times required for a given level of accuracy can be comparable for either image criterion.

Notable is that reconstruction quality, as characterized by criteria such as the point-spread function of a quantitative image, can be improved by optimization of the weight $\hat{v}(f)$.[21] Although the delay-and-sum reconstruction formula (20) depends on a frequency weight determined by the incident waveform $u(t)$, any desired weight $\hat{v}(f)$ can still be applied by preprocessing of the scattering data. That is, the inverse problem associated with an arbitrary incident waveform $w(t)$ (such as the impulse response of a particular electroacoustic transducer) can be transformed into the inverse problem associated with a desired waveform $u(t)$ by applying the deconvolution operation

$$[p_s(\boldsymbol{\theta},\boldsymbol{\alpha},t)]_{u(t)}=\mathbf{F}^{-1}\left[\frac{\hat{u}(f)}{\hat{w}(f)}\mathbf{F}[p_s(\boldsymbol{\theta},\boldsymbol{\alpha},t)]_{w(t)}\right], \tag{28}$$

where $\mathbf{F}$ denotes temporal Fourier transformation, to the measured scattering data. This operation transforms the measured data into the corresponding data that would be measured using an optimal incident pulse $u(t)$. For reasons of stability, the effective bandwidth of $\hat{u}(f)$ should be comparable to that of $\hat{w}(f)$ (as determined, for instance, by the noise floor of a given measurement).

The adaptive focusing implemented here employed a simple straight-ray approximation for wavefront aberration incurred in tissue. However, the principle of aberration correction by adaptive focusing should allow greater improvements to be gained using more complete distortion models. For example, the distortion caused by a strongly scattering medium can be accurately modeled using a full-wave computational method such as that of Ref. 33. In principle, appropriate deconvolution could be employed to remove the effects of the intervening medium for each incident-wave direction, measurement direction, and image location, so that an aberration-corrected reconstruction could then be performed by applying Eq. (20) to the corrected scattering data. In some cases, *a priori* information on the scattering medium may be exploited to improve the convergence of such adaptive focusing algorithms. This basic approach, in which a linearized reconstruction is performed on scattering data that has been transformed to remove higher-order scattering effects, is common to a number of existing nonlinear inverse scattering methods.[42]

The methods of aberration correction proposed here differ from most adaptive imaging methods for pulse-echo ultrasound (e.g., Refs. 12 and 16) because adaptive focusing is performed using a direct reconstruction of the medium rather than a simpler distortion estimate. Thus, aberration correction using quantitative imaging methods could be of great interest for pulse-echo systems such as current clinical scanners. However, the limited spatial-frequency information provided in pulse-echo mode[1,18] reduces the quality of quantitative images of this kind. One possible approach to increasing the spatial-frequency content of pulse-echo quantitative images could be to apply deconvolution to the scattered signals.[43–45] If such deconvolution methods could increase the spatial-frequency coverage sufficiently to obtain accurate (although possibly low-resolution) quantitative sound-speed maps, such maps could be employed directly for adaptive focusing in pulse-echo images.

VI. CONCLUSIONS

Two related approaches to aberration correction for quantitative ultrasonic imaging have been presented. These methods are based on approximate solutions to the linearized time-domain inverse scattering problem, implemented using adaptations of two previous time-domain diffraction tomography methods.[18,21] One approach, based on a delay-and-sum reconstruction formula, applies adaptive focusing based on estimates of the scattering medium. The other approach implements aberration correction by applying appropriate delays to a time-dependent reconstruction.

Numerical results show that each of the considered aberration correction approaches increases the parametric range of validity for time-domain diffraction tomography by about a factor of 2. The extended range of validity is sufficient to allow effective quantitative imaging of large-scale scattering media, such as the 75-mm breast model imaged here at 0.5 MHz. Adaptive focusing correction based on more complete scattering models could further increase this range of validity. Given sufficient *a priori* information on the unknown medium, the principle of focus correction may allow accurate quantitative images to be obtained for strongly-scattering media at larger scales and higher frequencies.

The approaches presented here may also be useful for aberration correction in pulse-echo imaging. If sufficiently broadband information can be extracted from pulse-echo scattering data, the time-domain diffraction tomography methods considered here may allow quantitative tissue characterization using clinically convenient measurement configurations. Quantitative maps obtained in this manner would also be useful as medium models for aberration correction in conventional B-scan and synthetic-aperture imaging.

ACKNOWLEDGMENTS

This work was supported by the Breast Cancer Research Program of the U.S. Army Medical Research and Materiel Command under Grant No. DAMD17-98-1-8141. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the U.S. Army. The author is grateful for helpful discussions with James F. Kelly, Adrian I. Nachman, Feng Lin, and Robert C. Waag.

[1] E. Wolf, ''Three-dimensional structure determination of semi-transparent objects from holographic data,'' Opt. Commun. **1**, 153–156 (1969).
[2] A. J. Devaney, ''Inversion formula for inverse scattering within the Born approximation,'' Opt. Lett. **1**, 111–112 (1982).
[3] D. T. Borup, S. A. Johnson, W. W. Kim, and M. J. Berggren, ''Nonperturbative diffraction tomography via Gauss-Newton iteration applied to the scattering integral equation,'' Ultrason. Imaging **14**, 69–85 (1992).
[4] S. Gutman and M. Klibanov, ''Two versions of quasi-Newton method for multidimensional inverse scattering problem,'' J. Comput. Acoust. **1**, 197–228 (1993).
[5] L. M. Hinkelman, D.-L. Liu, R. C. Waag, Q. Zhu, and B. D. Steinberg, ''Measurement and correction of ultrasonic pulse distortion produced by the human breast,'' J. Acoust. Soc. Am. **97**, 1958–1969 (1995).
[6] T. D. Mast, L. M. Hinkelman, M. J. Orr, V. W. Sparrow, and R. C. Waag, ''Simulation of ultrasonic pulse propagation through the abdominal wall,'' J. Acoust. Soc. Am. **102**, 1177–1190 (1997). [Erratum: **104**, 1124–1125 (1998).]
[7] J. Ylitalo, E. Alasaarela, and J. Koivukangas, ''Ultrasound holographic B-scan imaging,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **36**, 376–383 (1989).
[8] K. E. Thomenius, ''Evolution of ultrasound beamformers,'' Proc. IEEE Ultrason. Symp. **2**, 1615–1622 (1996).
[9] L. M. Hinkelman, T. D. Mast, L. A. Metlay, and R. C. Waag, ''The effect of abdominal wall morphology on ultrasonic pulse distortion. Part I: Measurements,'' J. Acoust. Soc. Am. **104**, 3635–3649 (1998).
[10] S. W. Smith, G. E. Trahey, and O. T. von Ramm, ''Phased array ultrasound imaging through planar tissue layers,'' Ultrasound Med. Biol. **12**(3), 229–243 (1986).
[11] G. Kossoff, D. A. Carpenter, D. E. Robinson, D. Ostry, and P. L. Ho, ''A sonographic technique to reduce beam distortion by curved interfaces,'' Ultrasound Med. Biol. **15**(4), 375–382 (1989).
[12] M. O'Donnell and S. W. Flax, ''Phase-aberration correction using signals from point reflectors and diffuse scatterers: Measurements,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**(6), 768–774 (1988).
[13] L. Nock, G. E. Trahey, and S. W. Smith, ''Phase aberration correction in medical ultrasound using speckle brightness as a quality factor,'' J. Acoust. Soc. Am. **85**, 1819–1833 (1989).
[14] D. Rachlin, ''Direct estimation of aberrating delays in pulse-echo imaging systems,'' J. Acoust. Soc. Am. **88**, 191–198 (1990).
[15] M. Hirama, O. Ikeda, and T. Sato, ''Adaptive ultrasonic array imaging through an inhomogeneous layer,'' J. Acoust. Soc. Am. **71**, 100–109 (1982).
[16] D.-L. Liu and R. C. Waag, ''Correction of ultrasonic wavefront distortion using backpropagation and a reference waveform method for time-shift compensation,'' J. Acoust. Soc. Am. **96**, 649–660 (1994).
[17] C. Dorme and M. Fink, ''Ultrasonic beam steering through inhomogeneous layers with a time reversal mirror,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**(1), 167–175 (1996).
[18] T. D. Mast, ''Wideband quantitative ultrasonic imaging by time-domain diffraction tomography,'' J. Acoust. Soc. Am. **106**, 3061–3071 (1999).
[19] T. D. Mast, F. Lin, and R. C. Waag, ''Time-domain ultrasound diffraction tomography,'' *1999 IEEE Ultrasonics Symposium Proceedings*, Vol. 2, pp. 1617–1620.
[20] M. Slaney, A. C. Kak, and L. E. Larsen, ''Limitations of imaging with first-order diffraction tomography,'' IEEE Trans. Microwave Theory Tech. **32**, 860–874 (1984).
[21] F. Lin, A. I. Nachman, and R. C. Waag, ''Quantitative imaging using a time-domain eigenfunction method,'' J. Acoust. Soc. Am. **108**, 899–912 (2000).
[22] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. I, Chap. 7.
[23] A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications*, 2nd ed. (Acoustical Society of America, Woodbury, NY, 1989), Chap. 9.
[24] G. Beylkin, ''The fundamental identity for iterated spherical means and the inversion formula for diffraction tomography and inverse scattering,'' J. Math. Phys. **24**, 1399–1400 (1983).
[25] C. Gasquet and P. Witomski, *Fourier Analysis and Applications: Filtering, Numerical Computation, Wavelets* (Springer-Verlag, New York, 1999).
[26] K. Kim and E. Wolf, ''Non-radiating monochromatic sources and their fields,'' Opt. Commun. **59**, 1–6 (1986).
[27] N. Bleistein and J. K. Cohen, ''Nonuniqueness in the inverse source problem in acoustics and electromagnetics,'' J. Math. Phys. **18**, 194–201 (1977).
[28] T. Melamed, Y. Ehrlich, and E. Heymann, ''Short-pulse inversion of inhomogeneous media: a time-domain diffraction tomography,'' Inverse Probl. **12**, 977–993 (1996).
[29] T. D. Mast, A. I. Nachman, and R. C. Waag, ''Focusing and imaging using eigenfunctions of the scattering operator,'' J. Acoust. Soc. Am. **102**, 715–725 (1997).
[30] V. A. Burov and O. D. Rumyantseva, ''Linearized inverse problem of scattering in monochromatic and pulse modes,'' Acoust. Phys. **40**, 34–42 (1996).
[31] T. T. Jansson, T. D. Mast, and R. C. Waag, ''Measurements of differential scattering cross section using a ring transducer,'' J. Acoust. Soc. Am. **103**, 3169–3179 (1998).
[32] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Chap. 8.
[33] T. D. Mast, L. P. Souriau, D.-L. Liu, M. Tabei, A. I. Nachman, and R. C. Waag, ''A *k*-space method for large-scale models of wave propagation in

tissue,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 341–354 (2001).
[34] T. D. Mast, ''Empirical relationships between acoustic parameters in human soft tissues,'' Acoust. Res. Lett. Online **1**, 37–42 (2000).
[35] J. L. Aroyan, ''Three-dimensional modeling of hearing in *Delphinus delphis*,'' J. Acoust. Soc. Am. **110**, 3305–3318 (2001).
[36] S. Wang, ''Finite-difference time-domain approach to underwater acoustic scattering problems,'' J. Acoust. Soc. Am. **99**, 1924–1931 (1996).
[37] M. Frigo and S. G. Johnson, ''FFTW: An adaptive software architecture for the FFT,'' *Proceedings of the ICASSP*, Vol. 3, pp. 1381–1384 (1998).
[38] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran* (Cambridge U.P., New York, 1986), Chap. 3.3.
[39] D. F. Rogers, *Procedural Elements for Computer Graphics* (McGraw–Hill, Boston, 1998), Chap. 2.
[40] A. E. Yagle and B. C. Levy, ''Layer-stripping solutions of multidimensional inverse scattering problems,'' J. Math. Phys. **27**, 1701–1710 (1986).
[41] Y. Chen, ''Inverse scattering via skin effect,'' Inverse Probl. **13**, 647–667 (1997).
[42] R. Snieder, ''A perturbative analysis of non-linear inversion,'' Geophys. J. Int. **101**, 545–556 (1990).
[43] J. A. Jensen and S. Leeman, ''Nonparametric estimation of ultrasound pulses,'' IEEE Trans. Biomed. Eng. **41**, 929–936 (1994).
[44] P. Lasaygues and J.-P. Lefebvre, ''Improvement of resolution in ultrasonic reflection tomography,'' The e-Journal of Nondestructive Testing **3**(8) (1998) [http://www.ndt.net/].
[45] O. Husby, T. Lie, T. Lango, J. Hokland, and H. Rue, ''Bayesian 2-D deconvolution: a model for diffuse ultrasound scattering,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 121–130 (2001).

Prediction of the generation of acoustic waves due to the penetration of pulsed microwaves in multilayer media

Emmanuel Guilliorit, Christophe Bacon,[a] and Bernard Hosten
Laboratoire de Mécanique Physique, Université Bordeaux 1, UMR CNRS 5469, 351 cours de la Libération, 33405 TALENCE Cedex, France

(Received 29 January 2002; accepted for publication 8 April 2002)

The acoustic wave generation in a body irradiated by a pulsed microwave is predicted theoretically. The irradiated body is a viscoelastic multilayer rod inserted into a waveguide or an uniformly irradiated viscoelastic plate. The model is based on Maxwell's equations, the heat equation, and thermoviscoelasticity theory. It is validated experimentally by means of four tests performed on three different specimens. The two first specimens are homogeneous rods used to evaluate and verify the mechanical and electromagnetic characteristics of two different materials. The third specimen is a composite rod made up of these materials. Two tests are performed with this specimen. The comparison between the experimental results and the theoretical computations leads to the validation of the theoretical model. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1482074]

PACS numbers: 43.20.Gp, 43.20.Mv, 43.20.Jr [YHB]

I. INTRODUCTION

A great number of papers deal with the generation of acoustic waves resulting from thermal expansion produced in an elastic body by sudden heating due to electromagnetic energy penetration (optical penetration of laser pulse or microwave irradiation).[1–14]

Hutchins has presented an extensive review of experimental and theoretical works in the case of laser-generated ultrasound.[1] In most studies concerned with ultrasound generation by laser, it is considered that the spatial profile of the temperature rise due to optical penetration is exponential and that the depth of the optical penetration is very short compared to the thickness of the illuminated body or compared to the thickness of the illuminated layer in the case of a multilayer body. Then, the acoustic waves are generated in a small region very close to the illuminated surface and the effects of the acoustic generation and of the acoustic propagation can be mathematically separated. This is true only when the penetration depth is small. If the optical penetration depth is significant compared to the considered propagation length, the effects of the acoustic generation and of the acoustic propagation cannot be separated (for instance, in the case of optically transparent materials).

This statement is indeed generally not valid in the case of polymers irradiated by microwaves. For many polymers, the microwave absorption is indeed very low and the penetration depth may be very large compared to the thickness of the irradiated body. Consequently, the acoustic wave generation may occur not only near the irradiated surface but also at the opposite surface or inside the specimen. This fact has been shown experimentally for polymer rods irradiated by microwaves.[10,11,14] An identical problem may occur when very thin layers are illuminated by short laser pulses. Then, it is difficult to separate the acoustic generation and the acoustic propagation mathematically.

In recent works, the problem of the acoustic wave generation in a viscoelastic body by the absorption of microwave irradiation has been treated in different cases.[10–14] The first works deal with the prediction of the acoustic waves generated in homogeneous viscoelastic rods by assuming empirically uniform, linear[11] or exponential[12] profile of the temperature rise. These cases have been extended to the case of an irradiated multilayer viscoelastic body[13] (plates or rods), where the temperature rise is assumed to be known (exponential or linear in each layer).

For homogeneous viscoelastic rods, a new theoretical model has been developed where the temperature rise is not supposed empirically but calculated from the knowledge of the incident electromagnetic wave and from the electromagnetic and the thermal characteristics of the materials.[14] In the present study, this prediction of the temperature rise and of the corresponding generated acoustic waves will be extended to the case of an irradiated multilayer viscoelastic body.

The theory is presented in Sec. II. Our aim in this section is to predict the acceleration or the velocity at the end of an irradiated rod embedded in a waveguide or at the surface of a uniformly irradiated plate. In Sec. III, the theory will be validated experimentally by means of four different tests performed on irradiated homogeneous or inhomogeneous rods. The inhomogeneous rod will be made up of two rods glued together.

II. THEORY

Consider an elastic or viscoelastic rod that can be divided into n segments #j ($z_j \leqslant z \leqslant z_{j+1}$ where z is the axial coordinate; see Fig. 1). The length of the segment #j is L_j. In each segment, the mechanical characteristics, the mass density ρ_j, and the cross-sectional area A are uniform. The co-

[a]Author to whom correspondence should be addressed. Laboratoire de Mécanique Physique, Université Bordeaux 1, UMR CNRS 5469, 351 cours de la Libération, 33405 TALENCE Cedex, France. Electronic mail: bacon@lmp.u-bordeaux.fr; Telephone: +33(0)5 57 96 22 72; fax: +33 (0)5 56 84 69 64.

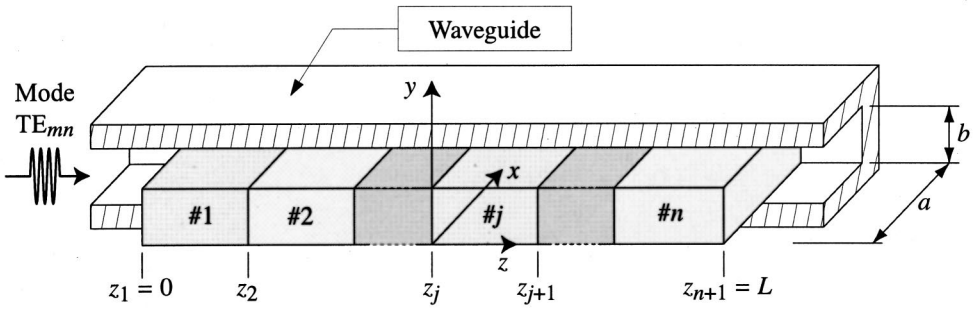


FIG. 1. Multilayer rod put in a rectangular waveguide.

ordinate at the first end of the body is $z_1=0$ and the coordinate at the second end is $z_n=L$. The rod is put in a rectangular waveguide of internal lateral dimensions a and b. There is enough play between the rod and the waveguide so that the rod can slide freely.

The free surface $z=0$ is irradiated by a time-gated microwave at frequency ν_0 in the TE$_{mn}$ electromagnetic mode. The pulse width is τ and the incident electromagnetic power (mean power) is $\overline{P^i}$. Owing to the electromagnetic absorption, a temperature rise $\theta(z,t)$ occurs inside the rod (t is the time). The temperature is not necessarily continuous because the heat conduction is neglected in the materials and at the interfaces. This assumption is valid because the test durations are very short while the effects of the heat conduction are very slow in comparison.

The case of an infinite electromagnetically isotropic multilayer plate can be treated simultaneously if this plate is irradiated uniformly on its first surface. Then, the lateral dimensions a and b can be chosen to be equal to one.

A. Determination of the temperature rise

To determine the temperature reached by the irradiated material, the instantaneous electromagnetic power must be known.

Maxwell's equations, in the absence of currents and charges, are

$$\nabla\cdot\tilde{\mathbf{E}}=0,\quad \nabla\cdot\tilde{\mathbf{H}}=0,\quad \nabla\times\tilde{\mathbf{E}}=-\mu_0 i\omega\tilde{\mathbf{H}},$$
$$\nabla\times\tilde{\mathbf{H}}=\epsilon_j^* i\omega\tilde{\mathbf{E}}, \tag{1}$$

where ω is the angular frequency related to the frequency ν by $\omega=2\pi\nu$, $\tilde{\mathbf{E}}$ is the Fourier transform of the electric field intensity $\mathbf{E}$, $\tilde{\mathbf{H}}$ is the Fourier transform of the magnetic field intensity $\mathbf{H}$, μ_0 is the magnetic permeability, and $\epsilon_j^*=\epsilon_j'-i\epsilon_j''$ is the complex dielectric permittivity of the material in segment #j.

First, consider that the fields are due to waves propagating in both positive and negative z directions in an infinite waveguide filled with an absorbing material:

$$\tilde{\mathbf{E}}(x,y,z,\omega)=\tilde{\mathbf{E}}^+(x,y,\omega)e^{-ik^*z}+\tilde{\mathbf{E}}^-(x,y,\omega)e^{ik^*z},$$
$$\tilde{\mathbf{H}}(x,y,z,\omega)=\tilde{\mathbf{H}}^+(x,y,\omega)e^{-ik^*z}+\tilde{\mathbf{H}}^-(x,y,\omega)e^{ik^*z}, \tag{2}$$

where $k^{*2}=(k'-ik'')^2$ is the complex electromagnetic wave number. The real part $k'(\omega)$ of k_z^* is an odd function of frequency; its imaginary part $-k''(\omega)$ is an even function. The instantaneous power on a section having cross-sectional area A is

$$P(z,t)=\int_A \mathbf{\Pi}\cdot\vec{e}_z\,dS, \tag{3}$$

where $\mathbf{\Pi}=\mathbf{E}\times\mathbf{H}$ is the Poynting vector. In the frequency domain the Fourier transform of the instantaneous power is written using the convolutions of the fields,

$$\tilde{P}(z,\omega)=\frac{1}{2\pi}\int_A(\tilde{E}_x*\tilde{H}_y-\tilde{E}_y*\tilde{H}_x)dS, \tag{4}$$

where $\tilde{E}_x*\tilde{H}_y$ stands for the convolution of $\tilde{E}_x$ and $\tilde{H}_y$, defined by

$$\tilde{E}_x*\tilde{H}_y=\int_{-\infty}^{+\infty}\tilde{E}_x(\Omega)\tilde{H}_y(\omega-\Omega)d\Omega. \tag{5}$$

Values of $\tilde{E}_y$ and $\tilde{H}_x$ can be deduced from values of $\tilde{E}_x$ and $\tilde{H}_y$ by using Maxwell's equations. Then, just the knowledge of $\tilde{E}_x$ and $\tilde{H}_y$ is required.

For $z<0$, the waveguide is filled with a semi-infinite medium called the *incident* medium (superscript i). For $z>L$ the waveguide is filled with a semi-infinite medium called the *transmitted* medium (superscript t). The continuity of the tangential components of the electric and magnetic fields at each interface gives a relation between the tangential components of fields at interface $z=0$ and these components at interface $z=L$. This relation is

$$\begin{bmatrix}\tilde{E}_t\\ \tilde{H}_t\end{bmatrix}=P\begin{bmatrix}\tilde{E}_i\\ \tilde{H}_i\end{bmatrix}=P^{(n)}\cdot P^{(n-1)}\cdot\ \cdots\ \cdot P^{(j)}\cdot\ \cdots\ \cdot P^{(1)}\begin{bmatrix}\tilde{E}_i\\ \tilde{H}_i\end{bmatrix}, \tag{6}$$

where $P^{(j)}$ is the transfer matrix of segment #j. This matrix must be calculated now. In segment #j, the tangential components $\tilde{E}_{jx}$ and $\tilde{H}_{jy}$ of fields propagating in the z direction are

$$\tilde{E}_{jx}(x,y,z,\omega)=a_x(x,y)\tilde{E}_{jx}(z,\omega)$$
$$=a_x(x,y)[\tilde{E}_{jx}^+(\omega)e^{-ik_j^*z}+\tilde{E}_{jx}^-(\omega)e^{ik_j^*z}],$$
$$\tilde{H}_{jy}(x,y,z,\omega)=a_x(x,y)\tilde{H}_{jy}(z,\omega) \tag{7}$$
$$=a_x(x,y)[\tilde{H}_{jy}^+(\omega)e^{-ik_j^*z}+\tilde{H}_{jy}^-(\omega)e^{ik_j^*z}],$$

where $a_x(x,y)$ is a function depending of the lateral dimensions of the irradiated body. The wave number k_j^* is the wave number of segment #j. If the length of segment #j is L_j, at interfaces (j) and $(j+1)$, the electric and magnetic fields are, respectively,

$$\tilde{E}_{jx}(\omega)=\tilde{E}^{+}_{jx}(\omega)+\tilde{E}^{-}_{jx}(\omega),$$
$$\tilde{H}_{jy}(\omega)=\tilde{H}^{+}_{jy}(\omega)+\tilde{H}^{-}_{jy}(\omega), \tag{8}$$
$$\tilde{E}_{(j+1)x}(\omega)=\tilde{E}^{+}_{jx}(\omega)e^{-ik^{*}_{j}L_j}+\tilde{E}^{-}_{jx}(\omega)e^{ik^{*}_{j}L_j},$$
$$\tilde{H}_{(j+1)y}(\omega)=\tilde{H}^{+}_{jy}(\omega)e^{-ik^{*}_{j}L_j}+\tilde{H}^{-}_{jy}(\omega)e^{ik^{*}_{j}L_j}. \tag{9}$$

Considering a TE mode ($E_z=0$), it can be demonstrated from the Maxwell's equations $\nabla\times\tilde{\mathbf{E}}=-\mu_0 i\omega\tilde{\mathbf{H}}$ that

$$\tilde{H}^{\pm}_{y}=\pm\frac{k^{*}(\omega)}{\omega\mu_0}\tilde{E}^{\pm}_{x}=\pm Y(\omega)\tilde{E}^{\pm}_{x}. \tag{10}$$

The quantity $Y(\omega)$ equals Y_j in segment #j. The introduction of Eq. (10) in Eqs. (8) and (9) leads to

$$\begin{bmatrix}\tilde{E}_{(j+1)x}(\omega)\\ \tilde{H}_{(j+1)y}(\omega)\end{bmatrix}=P^{(j)}\begin{bmatrix}\tilde{E}_{jx}(\omega)\\ \tilde{H}_{jy}(\omega)\end{bmatrix}$$
$$=\begin{bmatrix}\cos(k^{*}_{j}L_j) & -\dfrac{i}{Y_j}\sin(k^{*}_{j}L_j)\\ -iY_j\sin(k^{*}_{j}L_j) & \cos(k^{*}_{j}L_j)\end{bmatrix}$$
$$\times\begin{bmatrix}\tilde{E}_{jx}(\omega)\\ \tilde{H}_{jy}(\omega)\end{bmatrix}. \tag{11}$$

This equation allows us to calculate the transfer matrix P of the whole irradiated body by means of Eq. (6) and then the fields at each interface if the fields at the first interface are known. However, only the field $\tilde{E}^{+}_{ix}$ due to the incident electromagnetic wave is known (wave propagating in the positive direction in the *incident* medium). It can be written that

$$\begin{bmatrix}\tilde{E}^{+}_{tx}(\omega)\\ \tilde{H}^{+}_{ty}(\omega)\end{bmatrix}=P\begin{bmatrix}\tilde{E}_{ix}(\omega)\\ \tilde{H}_{iy}(\omega)\end{bmatrix}=P\begin{bmatrix}\tilde{E}^{+}_{ix}(\omega)+\tilde{E}^{-}_{ix}(\omega)\\ \tilde{H}^{+}_{iy}(\omega)+\tilde{H}^{-}_{iy}(\omega)\end{bmatrix}. \tag{12}$$

From Eq. (10), this equation becomes (Y_i and Y_t in the *incident* and *transmitted* media)

$$\begin{bmatrix}\tilde{E}^{+}_{tx}(\omega)\\ Y_i\tilde{E}^{+}_{tx}(\omega)\end{bmatrix}=P\begin{bmatrix}\tilde{E}^{+}_{ix}(\omega)+\tilde{E}^{-}_{ix}(\omega)\\ Y_i\tilde{E}^{+}_{ix}(\omega)-Y_i\tilde{E}^{-}_{ix}(\omega)\end{bmatrix}. \tag{13}$$

From this equation, the reflection coefficient $R(\omega)$ of the whole multilayer body is defined by

$$\tilde{E}^{-}_{ix}(\omega)=R(\omega)\tilde{E}^{+}_{ix}(\omega)$$
$$=\frac{(P_{21}+Y_iP_{22})-Y_t(P_{11}+Y_iP_{12})}{Y_t(P_{11}-Y_iP_{12})-(P_{21}-Y_iP_{22})}\tilde{E}^{+}_{ix}(\omega). \tag{14}$$

Then, it can be written that ($M^{(1)}$ is the identity matrix)

$$\begin{bmatrix}\tilde{E}_{jx}(\omega)\\ \tilde{H}_{jy}(\omega)\end{bmatrix}=P^{(j-1)}.\ldots.P^{(1)}\begin{bmatrix}\tilde{E}^{+}_{ix}(\omega)(1+R)\\ Y_i\tilde{E}^{+}_{ix}(\omega)(1-R)\end{bmatrix}$$
$$=M^{(j)}\begin{bmatrix}\tilde{E}^{+}_{ix}(\omega)(1+R)\\ Y_i\tilde{E}^{+}_{ix}(\omega)(1-R)\end{bmatrix}. \tag{15}$$

At this step, the tangential components of the electric and magnetic fields are known at each interface if the quantity $\tilde{E}^{+}_{ix}$ is known in the *incident* medium. In order to calculate these fields in a segment #j between two interfaces, Eq. (7) shows that it is necessary to calculate the quantities $\tilde{E}^{+}_{jx}$ and $\tilde{E}^{-}_{jx}$. From Eqs. (8) and (15), it can be derived that

$$\tilde{E}^{+}_{jx}(\omega)=t_j(\omega)\tilde{E}^{+}_{ix}(\omega)$$
$$=\frac{1}{2}\Bigg\{[(M^{(j)}_{11}+Y_iM^{(j)}_{12})+R(M^{(j)}_{11}-Y_iM^{(j)}_{12})]$$
$$+\frac{1}{Y_j}[(M^{(j)}_{21}+Y_iM^{(j)}_{22})$$
$$+R(M^{(j)}_{21}-Y_iM^{(j)}_{22})]\Bigg\}\tilde{E}^{+}_{ix}(\omega), \tag{16}$$

$$\tilde{E}^{-}_{jx}(\omega)=r_j(\omega)\tilde{E}^{+}_{ix}(\omega)$$
$$=\frac{1}{2}\Bigg\{[(M^{(j)}_{11}+Y_iM^{(j)}_{12})+R(M^{(j)}_{11}-Y_iM^{(j)}_{12})]$$
$$-\frac{1}{Y_j}[(M^{(j)}_{21}+Y_iM^{(j)}_{22})$$
$$+R(M^{(j)}_{21}-Y_iM^{(j)}_{22})]\Bigg\}\tilde{E}^{+}_{ix}(\omega). \tag{17}$$

Thus, if $z=0$ at interface (j), the fields in a segment #j between two interfaces are given by

$$\tilde{E}_{jx}(x,y,z,\omega)=a_x(x,y)\tilde{E}^{+}_{ix}(\omega)[t_j(\omega)e^{-ik^{*}_{j}z}$$
$$+r_j(\omega)e^{ik^{*}_{j}z}],$$
$$\tilde{H}_{jy}(x,y,z,\omega)=a_x(x,y)\tilde{E}^{+}_{ix}(\omega)Y_j(\omega)[t_j(\omega)e^{-ik^{*}_{j}z} \tag{18}$$
$$-r_j(\omega)e^{ik^{*}_{j}z}].$$

To simplify, it is more convenient to define the functions $M_j(z,\omega)$ and $N_j(z,\omega)$ as

$$M_j(z,\omega)=t_j(\omega)\left[e^{-ik^{*}_{j}z}+\frac{r_j(\omega)}{t_j(\omega)}e^{ik^{*}_{j}z}\right] \tag{19}$$

and

$$N_j(z,\omega)=t_j(\omega)\left[e^{-ik^{*}_{j}z}-\frac{r_j(\omega)}{t_j(\omega)}e^{ik^{*}_{j}z}\right]. \tag{20}$$

We can notice that

$$\frac{\partial M_j(z,\omega)}{\partial z}=-ik^{*}_{j}N_j(z,\omega)$$
and
$$\frac{\partial N_j(z,\omega)}{\partial z}=-ik^{*}_{j}M_j(z,\omega). \tag{21}$$

Then, the fields $\tilde{E}_{jx}$ and $\tilde{H}_{jy}$ are

$$\tilde{E}_{jx}(x,y,z,\omega)=a_x(x,y)M_j(z,\omega)\tilde{E}^{+}_{ix}(\omega),$$
$$\tilde{H}_{jy}(x,y,z,\omega)=a_x(x,y)N_j(z,\omega)Y_j(\omega)\tilde{E}^{+}_{ix}(\omega). \tag{22}$$

In the same way, the fields $\tilde{E}_{jy}$ and $\tilde{H}_{jx}$ can be calculated. We can obtain

$$\tilde{E}_{jy}(x,y,z,\omega)=a_y(x,y)M_j(z,\omega)\tilde{E}^+_{ix}(\omega),$$
$$\tilde{H}_{jx}(x,y,z,\omega)=-a_y(x,y)N_j(z,\omega)Y_j(\omega)\tilde{E}^+_{ix}(\omega). \tag{23}$$

The functions $k_j^*(\omega)$, $a_x(x,y)$, $a_y(x,y)$ can be calculated from Maxwell's equations (1). In the case of an electromagnetic propagation in a waveguide filled by a rod having lateral dimensions a and b in the x and y directions, it can be demonstrated that these functions are given by

$$k_j^{*2}=\mu_0\epsilon_j^*\omega^2-(k_x^2+k_y^2)=(k_j'-ik_j'')^2, \tag{24}$$

$$a_x(x,y)=-\frac{k_y}{k_x}c_xs_y \quad \text{and} \quad a_y(x,y)=s_xc_y, \tag{25}$$

where $k_x=m\pi/a$ and $k_y=n\pi/b$ for the TE_{mn} mode, $c_x=\cos k_x x$, $c_y=\cos k_y y$, $s_x=\sin k_x x$, $s_y=\sin k_y y$.

In the case of a uniformly irradiated plate (electromagnetic TEM plane wave, infinite lateral dimensions), where the electric field is parallel to the x axis, these quantities become

$$k_j^{*2}=\mu_0\epsilon_j^*\omega^2, \tag{26}$$

$$a_x(x,y)=1 \quad \text{and} \quad a_y(x,y)=0. \tag{27}$$

In this case, the electromagnetic power and other physical quantities will be calculated for a unit surface.

The introduction of Eq. (23) in Eq. (4) gives the power as

$$\tilde{P}(z,\omega)=\frac{1}{2\pi}\int_A(a_x^2+a_y^2)[M_j(z,\omega)\tilde{E}^+_{ix}(\omega)]*[N_j(z,\omega)Y_j(\omega)\tilde{E}^+_{ix}(\omega)]dS, \tag{28}$$

or

$$\tilde{P}(z,\omega)=\frac{1}{2\pi}f_{xy}[M_j(z,\omega)\tilde{E}^+_{ix}(\omega)]*[N_j(z,\omega)Y_j(\omega)\tilde{E}^+_{ix}(\omega)]. \tag{29}$$

with

$$f_{xy}=\int_A(a_x^2+a_y^2)dS. \tag{30}$$

For the determination of the temperature rise, it is necessary to know the gradient of the power in the propagation direction. This gradient is

$$\frac{\partial\tilde{P}_j(z,\omega)}{\partial z}=\frac{1}{2\pi}f_{xy}\left\{\left[\frac{\partial M_j(z,\omega)}{\partial z}\tilde{E}^+_{ix}(\omega)\right]*[N_j(z,\omega)Y_j(\omega)\tilde{E}^+_{ix}(\omega)]+[M_j(z,\omega)\tilde{E}^+_{ix}(\omega)]*\left[\frac{\partial N_j(z,\omega)}{\partial z}Y_j(\omega)\tilde{E}^+_{ix}(\omega)\right]\right\}. \tag{31}$$

Equations (21) lead to

$$\frac{\partial\tilde{P}_j(z,\omega)}{\partial z}=-\frac{i}{2\pi}f_{xy}\{[k_j^*N_j(z,\omega)\tilde{E}^+_{ix}(\omega)]*[N_j(z,\omega)Y_j(\omega)\tilde{E}^+_{ix}(\omega)]+[M_j(z,\omega)\tilde{E}^+_{ix}(\omega)]*[k_j^*M_j(z,\omega)Y_j(\omega)\tilde{E}^+_{ix}(\omega)]\}. \tag{32}$$

For an incident sinusoidal wave at angular frequency $\omega_0=2\pi\nu_0$, the field can be written as

$$\tilde{E}^+_{ix}(\omega)=E^+_{ix}\frac{\pi}{i}[\delta(\omega-\omega_0)-\delta(\omega+\omega_0)], \tag{33}$$

where $\delta(\omega)$ is the Dirac function. The assumption of a continuous sinusoidal wave is valid in our case because the duration τ of the emitted pulses will be very large compared to the period T_0 of one oscillation. For instance, one pulse of 1 μs contains 10 000 oscillations at a frequency of about 10 GHz. For one period T_0, it can be demonstrated that the mean values of the power and its derivative are

$$\overline{P_j}(z)=\frac{1}{T_0}\int_0^{T_0}P(z,t)dt=\frac{f_{xy}}{4\mu_0\omega_0}(E^+_{ix})^2[M_j(z,-\omega_0)k_j^*(\omega_0)N_j(z,\omega_0)-M_j(z,\omega_0)k_j^*(-\omega_0)N_j(z,-\omega_0)], \tag{34}$$

$$\frac{\partial\overline{P_j}(z)}{\partial z}=\frac{1}{T_0}\int_0^{T_0}\frac{\partial P(z,t)}{\partial z}dt=\frac{if_{xy}}{4\mu_0\omega_0}(E^+_{ix})^2M_j(z,\omega_0)M_j(z,-\omega_0)\times[k_j^{*2}(-\omega_0)-k_j^{*2}(\omega_0)]. \tag{35}$$

Generally, the incident mean power in the incident medium is known but not the incident electric field. The incident medium is considered to be a nonabsorbing medium (vacuum). The mean power transported by the incident wave in a nonabsorbing incident medium $[k_i''(\omega)=0]$ is calculated by writing $t(\omega)=1$ and $r(\omega)=0$ in Eq. (19). We obtain from Eq. (34),

$$\overline{P_i}=\frac{k_i'(\omega_0)}{2\mu_0\omega_0}f_{xy}(E^+_{ix})^2. \tag{36}$$

It is assumed that the microwave sinusoidal train is emitted during a very short time inducing a short thermal pulse and that the depth of energy deposition is large compared to any relevant thermal diffusion length. Then, heat conduction can be neglected. Then, the heat equation is

$$\rho_jC_j\frac{\partial\theta_j}{\partial t}=-\frac{1}{A}\frac{\partial P_j}{\partial z}, \tag{37}$$

where θ is the temperature rise, and C_j is the specific heat. As a consequence the temperature rise after one period is given by

$$\theta_j(z,T_0) = -\frac{1}{\rho_j C_j A}\int_0^{T_0}\frac{\partial P_j(z,t)}{\partial z}dt$$

$$= -\frac{2\pi}{\rho_j C_j A\omega_0}\frac{\partial \overline{P}_j(z)}{\partial z}. \tag{38}$$

From Eqs. (35) and (36), this temperature rise becomes

$$\theta_j(z,T_0) = \frac{4\pi \overline{P}_i k_j'(\omega_0)k_j''(\omega_0)}{\rho_j C_j A\omega_0 k_i'(\omega_0)} M_j(z,\omega_0)M_j(z,-\omega_0). \tag{39}$$

Since the period T_0 is generally very short compared to the pulse width τ, the temperature rise is

$$\theta_j(z,t) = \frac{\theta_j(z,T_0)}{T_0}\tau\cdot r(t) = \frac{\theta_j(z,T_0)\omega_0}{2\pi}\tau\cdot r(t), \tag{40}$$

where the time function $r(t)$ for one pulse is

$$r(t) = \begin{cases} 0, & \text{for } t\leqslant 0, \\ t/\tau, & \text{for } 0<t<\tau, \\ 1, & \text{for } t\geqslant \tau. \end{cases} \tag{41}$$

Then, it is convenient to write the Fourier transform of the temperature rise in the following way:

$$\tilde{\theta}_j(z,\omega) = \tilde{r}(\omega)K_j\sum_{k=1}^{4} C_{jk}e^{a_{jk}z}, \tag{42}$$

where $\tilde{r}(\omega)$ is the Fourier transform of the function $r(t)$ defined in Eq. (41),

$$K_j = \frac{2\overline{P}_i \tau k_j'(\omega_0)k_j''(\omega_0)}{\rho_j C_j A k_i'(\omega_0)} t_j(\omega_0)t_j(-\omega_0) \tag{43}$$

and

$$\begin{aligned} &C_{j1}=1, \quad a_{j1}=-2k_j''(\omega_0), \\ &C_{j2}=r_j(-\omega_0)/t_j(-\omega_0), \quad a_{j2}=-2ik_j'(\omega_0), \\ &C_{j3}=r_j(\omega_0)/t_j(\omega_0), \quad a_{j3}=-a_{j2}, \\ &C_{j4}=\frac{r_j(\omega_0)r_j(-\omega_0)}{t_j(\omega_0)t_j(-\omega_0)}, \quad a_{j4}=-a_{j1}. \end{aligned} \tag{44}$$

B. Determination of the acoustic waves

Two one-dimensional problems can be treated together: the case of a uniaxial strain state and the case of a uniaxial stress state. A uniaxial strain state concerns a problem where the lateral dimensions are important compared to the thickness of the body (uniformly irradiated plate). In this case, we can choose $A=1$ for any segment. It can be noticed that it is easy to irradiate a large surface uniformly by means of microwaves. The stress state can be considered uniaxial in a rod when the lateral dimensions are much smaller than the wavelengths (irradiated rod). In both cases, the Fourier transforms, $\tilde{\sigma}(z,\omega)$ and $\tilde{u}(z,\omega)$, of the normal stress $\sigma(z,t)$ and the axial displacement $u(z,t)$ in segment #j, respectively, are related by

$$\frac{\partial}{\partial z}\tilde{\sigma}(z,\omega) = -\rho_j\omega^2\tilde{u}(z,\omega), \tag{45}$$

where the angular frequency ω is related to the frequency ν by $\omega=2\pi\nu$. In the case of a uniaxial strain state, the linear viscoelastic constitutive law of the segment #j can be expressed in the frequency domain as

$$\tilde{\sigma}(z,\omega) = C_{11}^{*j}\tilde{\epsilon}(z,\omega) - (C_{11}^{*j}\alpha_1^j + C_{12}^{*j}\alpha_2^j + C_{13}^{*j}\alpha_3^j)\tilde{\theta}(z,\omega), \tag{46}$$

where C_{1k}^{*j} are the components of the complex stiffness matrix in segment #j (the index 1 denotes the orthotropic axis in the z direction, 2 and 3 the two other orthotropic axes), α_k^j are the coefficients of linear thermal expansion in the orthotropic directions, $\tilde{\epsilon}(z,\omega)$ and $\tilde{\theta}(z,\omega)$ the Fourier transforms of the longitudinal strain and the temperature rise above the initial temperature, respectively. In the case of a uniaxial stress state, this relation becomes

$$\tilde{\sigma}(z,\omega) = E_j^*\tilde{\epsilon}(z,\omega) - E_j^*\alpha_1^j\tilde{\theta}(z,\omega), \tag{47}$$

where E_j^* is the complex Young's modulus in the z direction. To facilitate the analysis, the temperature distribution will be assumed to be independent of the mechanical state of the material.

Both uniaxial problems can be treated together by introducing the complex slowness $S_j^*=S_j'-iS_j''$ and the expansion coefficient β_j, defined by the following: for a uniaxial strain state,

$$S_j^{*2} = \frac{\rho_j}{C_{11}^{*j}} \quad \text{and} \quad \beta_j = \frac{C_{11}^{*j}\alpha_1^j + C_{12}^{*j}\alpha_2^j + C_{13}^{*j}\alpha_3^j}{C_{11}^{*j}}; \tag{48}$$

for a uniaxial stress state,

$$S_j^{*2} = \frac{\rho_j}{E_j^*} \quad \text{and} \quad \beta_j = \alpha_1^j. \tag{49}$$

In the hypothesis of small strains, the longitudinal strain is related to the axial displacement by $\epsilon(z,t)=\partial u(z,t)/\partial z$. Consequently, it can be found[13] that

$$\frac{\partial^2\tilde{\epsilon}(z,\omega)}{\partial z^2} + \omega^2 S_j^{*2}\tilde{\epsilon}(z,\omega) - \beta_j\frac{\partial^2\tilde{\theta}(z,\omega)}{\partial z^2} = 0. \tag{50}$$

For an exponential profile of the temperature rise, the solution of this equation is given in Ref. 13. In the following, we introduce a state vector made of the Fourier transforms of the normal forces and the accelerations. The acceleration is chosen here since the generation mechanism produces a continuous component in the velocity and the displacement, contrary to the acceleration that is limited in time. It is demonstrated that the state vectors at interfaces (j) and ($j+1$) are related by

$$\begin{bmatrix} \tilde{F}_{j+1} \\ \tilde{a}_{j+1} \end{bmatrix} = P(j)\begin{bmatrix} \tilde{F}_j \\ \tilde{a}_j \end{bmatrix} + G^{(j)}. \tag{51}$$

In this case, the matrix $P^{(j)}$ is the acoustic transfer matrix in segment #j. This matrix is[13]

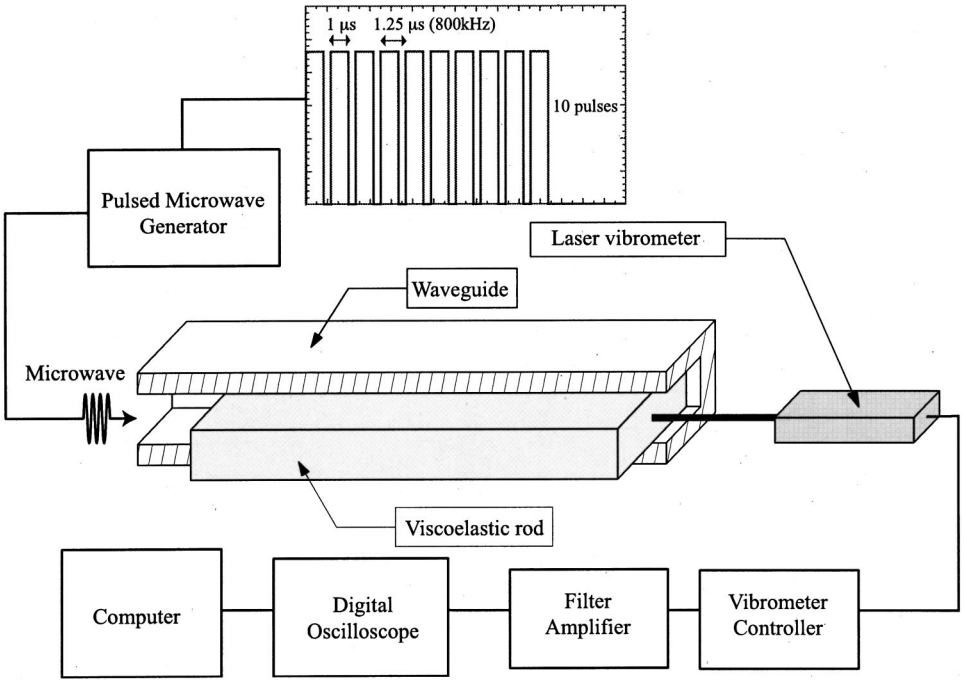


FIG. 2. Experimental setup (cutout of the waveguide).

$$P^{(j)}=\begin{bmatrix} c & \dfrac{\rho_j A}{\omega S_j^*}s \\ -\dfrac{\omega S_j^*}{\rho_j A}s & c \end{bmatrix}, \tag{52}$$

where $c=\cos(\omega S_j^* L_j)$ and $s=\sin(\omega S_j^* L_j)$. For the temperature profile given by Eqs. (42), (43) and (44), the generation vector $G^{(j)}$ of segment #j can be deduced from the exponential solution found in Ref. 13. It is found that

$$G^{(j)}=\beta_j \tilde{r} K_j \omega^2 \times \begin{bmatrix} \dfrac{\rho_j A}{\omega S_j^*}\displaystyle\sum_{k=1}^{4}\frac{C_{jk}}{a_{jk}^2+\omega^2 S_j^{*2}}(\omega S_j^* c + a_{jk}s - \omega S_j^* e^{a_{jk}L_j}) \\ \displaystyle\sum_{k=1}^{4}\frac{C_{jk}}{a_{jk}^2+\omega^2 S_j^{*2}}(-\omega S_j^* s + a_{jk}c - a_{jk}e^{a_{jk}L}) \end{bmatrix}. \tag{53}$$

Thanks to the continuity of the normal force and the acceleration, it can be shown[13] that

$$\begin{bmatrix} \tilde{F}_j \\ \tilde{a}_j \end{bmatrix} = S^{(j)}\begin{bmatrix} \tilde{F}_1 \\ \tilde{a}_1 \end{bmatrix} + T^{(j)}, \quad j \geqslant 2, \tag{54}$$

where

$$S^{(j)}=P^{(j-1)}S^{(j-1)}, \quad S^{(2)}=P^{(1)}, \qquad T^{(j)}=P^{(j-1)}T^{(j-1)}+G^{(j-1)}, \quad T^{(2)}=G^{(1)}. \tag{55}$$

Consequently, the normal force and the acceleration at both ends of the irradiated body are linked together by the matrix $S^{(n)}$ and the vector $T^{(n)}$. Then, the determination of the acceleration at $z=L$ (for instance) involves the knowledge of the boundary conditions. The end at $z=0$ is free. Consequently, the normal force at $z=0$ must be zero, i.e. $\tilde{F}_1=0$. For $z>L$, it is considered that there is no temperature rise. The function $H_L(\omega)=\tilde{F}(L,\omega)/\tilde{a}(L,\omega)$ is supposed to be known (for instance, in the case of a small accelerometer having mass m_a and mechanical impedance Z_a, $H_L=iZ_a/\omega$ with $Z_a=im_a\omega$;[13] if a contactless measurement method as a laser vibrometer is used, this function becomes $H_L(\omega)=0$). With the help of these boundary conditions and Eq. (54), the acceleration at $z=L$ is

$$\tilde{a}(L,\omega)=\tilde{a}_n=\frac{S_{22}^{(n)}T_1^{(n)}-S_{12}^{(n)}T_2^{(n)}}{H_L S_{22}^{(n)}-S_{12}^{(n)}}. \tag{56}$$

The Fourier transform $\tilde{v}(L,\omega)$ of the particle velocity $v(L,t)$ at the end $z=L$ is given by

$$\tilde{v}(L,\omega)=\frac{\tilde{a}(L,\omega)}{i\omega}. \tag{57}$$

Of course, this equation is not valid for the zero frequency. To obtain the temporal representation of the particle velocity $v(L,t)$ in the time domain, the acceleration $a(L,t)$ must be integrated numerically.

III. EXPERIMENTAL VERIFICATION

A. Experimental setup

The experimental setup is shown in Fig. 2 with a cutout of the waveguide in order to show the irradiated rod.

The microwave tonebursts are produced by a time-gated electromagnetic generator at 9.41 GHz. The maximum power of the incident electromagnetic wave is 5.5 kW, and the maximum pulse width is 1 μs. The wave is delivered by a standard waveguide plumbing (of cross section 23 $\times$10 mm) that permits only the fundamental waveguide mode TE_{10} to propagate. To obtain the maximum power from the generator, a hybrid tee impedance adaptor is inserted into the circuit and adjusted to achieve a minimum voltage standing wave ratio (VSWR), as measured by a Watt

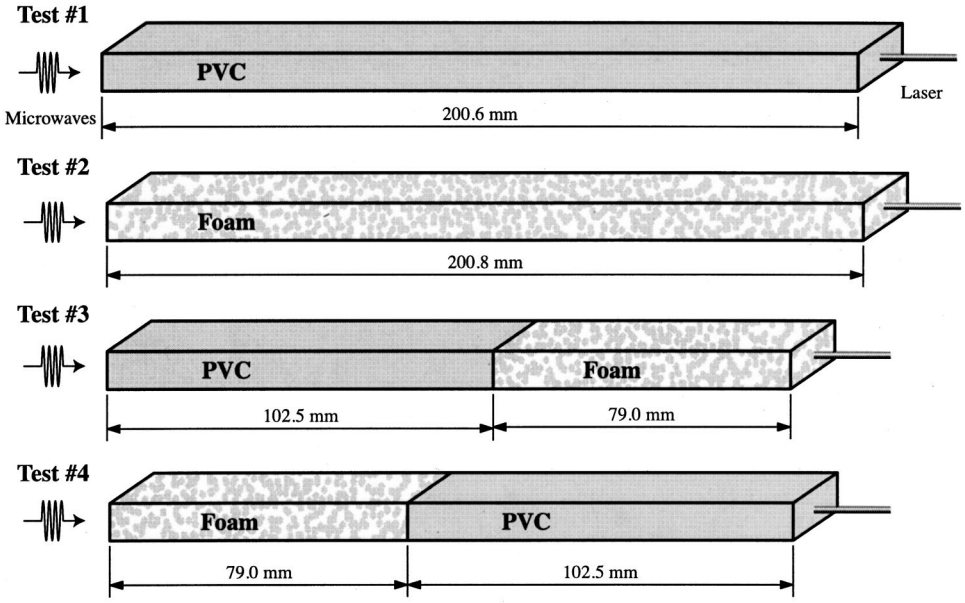


FIG. 3. Four performed tests.

meter. In order to increase the amplitude of the measured signal, a burst of 10 pulses of 1 μs are emitted with a repetition frequency of 800 kHz (see Fig. 2).

The rod sample is held horizontally inside the waveguide. The cross section of the rod is almost identical to the cross section of the waveguide, but there is enough play between them so that the specimen can slide freely. A laser vibrometer measures the velocity at the end of the irradiated rod. This laser vibrometer is made up of a sensor head (Polytec OFV-353) and a controller (Polytec OFV-3001) equipped with a velocity decoder (Polytec OVD-02). The measurement range used in our tests is 5 mm/s/V. Then the frequency range lies between 0.5 Hz and 250 kHz with a resolution of 0.5 μm/s. The signal from the controller is filtered by a filter amplifier (Electronic Instruments 3627) between about 1 and 20 kHz. The velocity signals are recorded on a digital oscilloscope (Lecroy 9310) that is, in turn, linked to a computerized data acquisition system. The signals are averaged in the oscilloscope for 50–300 sweeps in order to eliminate the electronic noise and the environmental vibrations. Then, the signals are transferred to a computer for further processing.

B. Results

In order to validate the theory presented in the previous section, four tests have been performed. These tests use three different samples made up with PVC and/or a rigid foam (see Fig. 3). These two materials have been chosen because of their different mechanical and electromagnetic characteristics.

Tests #1 and #2 are performed with homogeneous polymeric (PVC) and foam rod specimens in order to obtain or verify the mechanical and electromagnetic characteristics of these materials. The theory will be validated by predicting the mechanical responses of the irradiated composite specimen used in tests #3 and #4 and by comparing these theoretical responses to the experimental ones. For the calculations, characteristics obtained from tests #1 and #2 will be used. For each test, the rod specimen is irradiated at its left end and the velocity is measured at its right end.

In test #1, the PVC rod specimen is used. The PVC density equals 1380 kg/m^3. The length of the specimen is 200.6 mm. *A priori*, its specific heat, its coefficient of linear thermal expansion, its viscoelastic modulus (or its complex slowness), its complex electromagnetic permittivity are unknown. The complex slowness and the complex electromagnetic relative permittivity can be evaluated experimentally by using the method presented in Ref. 14. The method consists of measuring the oscillations in the particle acceleration or velocity detected at the end of the rod that are induced by variations in the polymer rod length. The following values have been obtained for the PVC material by means of this method: $S'=605$ μs/m, $S''=14$ μs/m, $\epsilon_r'=2.67$, $\epsilon_r''=0.02$. The corresponding Young's modulus is $E^*=(3.76+i0.17)$ GPa approximately. Once these characteristics have been obtained, the shape of the velocity spectrum at the end of the irradiated PVC specimen can be predicted theoretically. Then, by comparing the amplitudes of the experimental and the theoretical velocity spectra, the coefficient $\beta\overline{P_i}/C$ can be evaluated. The value obtained for the PVC is $\beta\overline{P_i}/C=4.1\ 10^{-4}$ kg. The knowledge of this coefficient, of the complex slowness and of the complex relative permittivity permits the complete computation of the theoretical velocity at the end of the irradiated PVC specimen (test #1). The theoretical velocity spectrum is compared with the experimental one in Fig. 4 which shows a very good agreement.

The product of the maximum temperature rise after one pulse by the coefficient of linear thermal expansion β can be also calculated. The result is presented in Fig. 5. The temperature rise can be calculated only if the coefficient of linear thermal expansion β is known. If we suppose that this coefficient is about $10^{-4}/°$C, it can be noticed that the maximum temperature rise after one pulse of 1 μs is about 3.10^{-5} °C.

In test #2, the foam rod specimen is used. The foam density equals 150 kg/m^3. The length of the specimen is

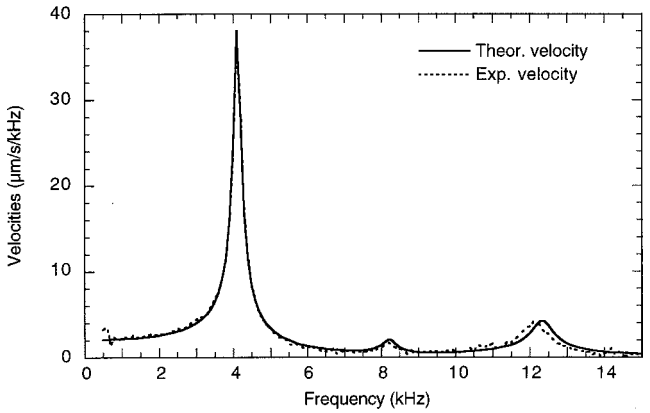


FIG. 4. Theoretical and experimental velocity spectra for the PVC specimen (test #1).

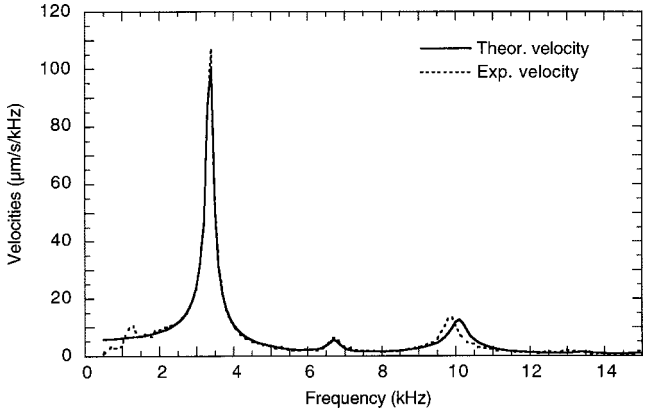


FIG. 6. Theoretical and experimental velocity spectra for the foam specimen (test #2).

200.8 mm. Contrary to the PVC specimen, the real part of the relative permittivity of the foam was given by the supplier. Its value is $\epsilon_r' = 1.25$ approximately. Then, the imaginary part of the relative permittivity and the complex slowness of the foam can be deduced from the experimental test #2 by using a numerical minimization procedure between the experimental velocity spectrum and the theoretical one. The minimization procedure allows us to find an optimal set of parameters $\mathbf{p}$ that provides a best fit of the theoretical spectrum to the experimental one. The function we minimize is given by

$$J(\mathbf{p}) = \sum_{\nu} \left| \frac{|\tilde{v}_{\exp}(\nu)|}{\Sigma_\nu |\tilde{v}_{\exp}(\nu)|} - \frac{|\tilde{v}(\nu,\mathbf{p})|}{\Sigma_\nu |\tilde{v}(\nu,\mathbf{p})|} \right|, \tag{58}$$

where $|\tilde{v}_{\exp}(\nu)|$ and $|\tilde{v}(\nu,\mathbf{p})|$ are the moduli of the Fourier transforms of the experimental and the theoretical velocities, respectively, and ν is the frequency. The set of adjustable parameters is $\mathbf{p} = (S', S'', \epsilon_r'')$. The following values have been obtained for the foam material by means of this method: $S' = 740\ \mu$s/m, $S'' = 16\ \mu$s/m, $\epsilon_r'' = 0.01$. The corresponding Young's modulus is $E^* = (0.27 + i0.01)$ GPa approximately. As explained above, from the comparison of the amplitudes of the experimental and the theoretical velocity spectra, the coefficient $\beta \overline{P_i}/C$ can be evaluated from test #2. For the foam, we obtain an experimental value equal to $\beta \overline{P_i}/C = 1.2\ 10^{-4}$ kg. Then, the theoretical and the experimental velocity spectra calculated or measured at the end of the irradiated foam are compared in Fig. 6.

In Fig. 4 and Fig. 6, the shift between the frequencies of the third peaks of the experimental and the theoretical velocities may be caused by the fact that it is assumed that the real part of the slowness does not depend on the frequency. This is an approximation.

The product of the maximum temperature rise after one pulse by the coefficient of linear thermal expansion β can be also calculated for the foam rod specimen. The result is presented in Fig. 7.

Once the mechanical and the electromagnetic characteristics of both PVC and foam materials have been evaluated and validated by means of tests #1 and #2, the theoretical equations developed in Sec. II permit the prediction of the response of any irradiated inhomogeneous rod made up of these both materials. For instance, the product of the maximum temperature rise by the coefficient of linear thermal expansion β can be predicted for the specimen used in tests #3 and #4. This specimen is made by gluing two rods made up of PVC and foam together. The theoretical results are shown in Fig. 8 and Fig. 9. The temperature rise cannot be calculated when the coefficients of linear thermal expansion β are unknown.

It is noticeable that the results shown in Fig. 8 and Fig.

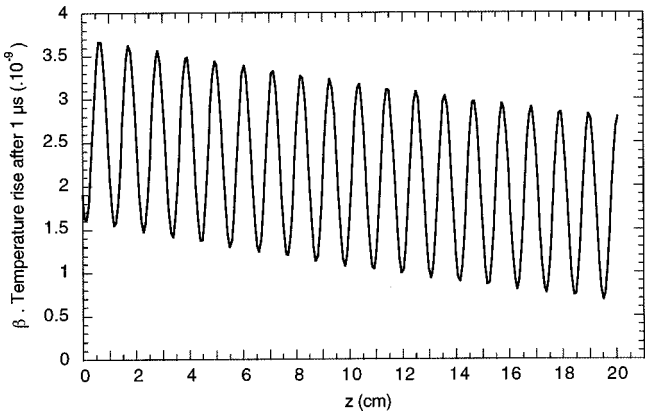


FIG. 5. Product of the maximum temperature rise after one pulse by the coefficient of linear thermal expansion β for the PVC specimen (test #1).

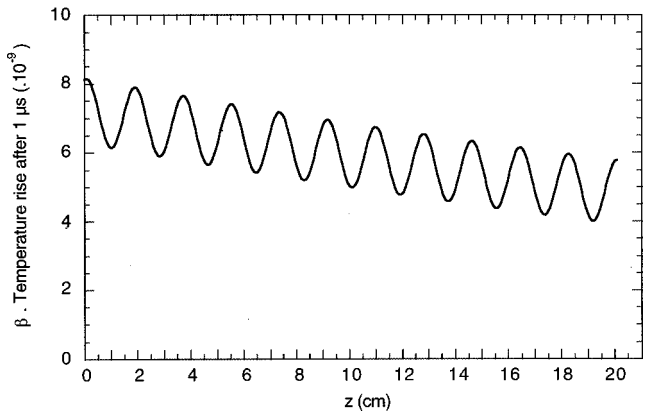


FIG. 7. Product of the maximum temperature rise after one pulse by the coefficient of linear thermal expansion β for the foam specimen (test #2).

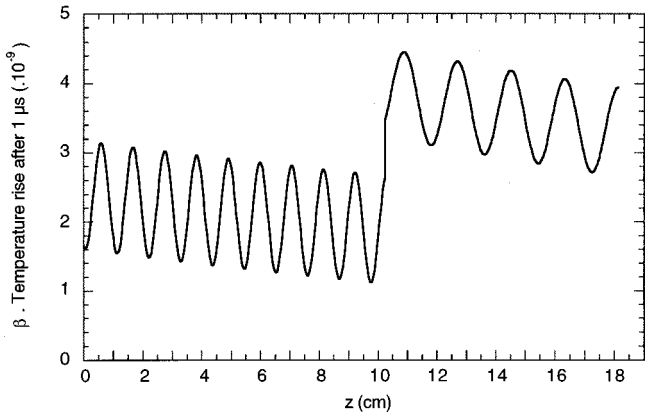


FIG. 8. Product of the maximum temperature rise after one pulse by the coefficient of linear thermal expansion β for the specimen PVC/foam (test #3).

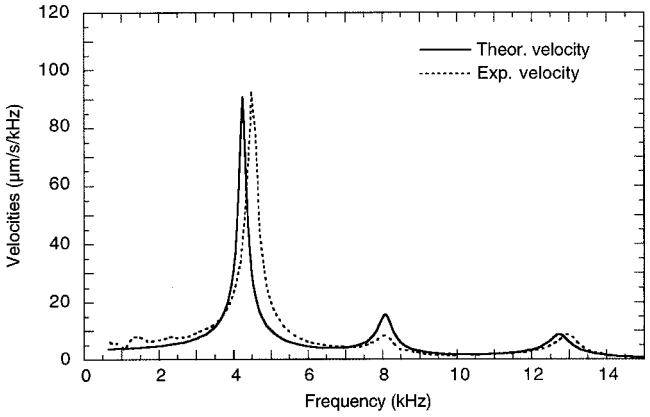


FIG. 10. Theoretical and experimental velocity spectra for the PVC/foam specimen (test #3).

9 are very different although the specimen used is the same one for both calculations. The specimen has just been turned. This fact was expected since the electromagnetic properties of both materials are quite different. In test #3, the electromagnetic wave is first partially absorbed by the PVC rod before propagating in the foam rod. In test #4, the electromagnetic wave is first partially absorbed by the foam rod.

The amplitude of the temperature rise is so small that it is difficult to measure. Thus, the theoretical results presented in Fig. 8 and Fig. 9 cannot be used to validate the theory. However, the particle velocities at the end of the specimen used in tests #3 and #4 can be measured and compared to the corresponding predicted velocities. The results for the velocity spectra are presented in Fig. 10 and Fig. 11.

It is noticeable that the spectra for tests #3 and #4 are very different. This was expected since it has been shown in Fig. 8 and Fig. 9 that the temperature rises in both tests are very different. For each test, the shapes of the predicted spectra are very close to the shape of the experimental ones. The small gaps between the theoretical and the experimental frequencies of the peaks in the spectra may be caused by the presence of the glue between the foam and the PVC rods. Due to its porosity, this foam is indeed impregnated by the glue close to the end in contact with the PVC rod. The amplitude of the velocity spectrum is predicted with good accuracy in the case of test #3. For test #4, the discrepancy between the amplitudes of the theoretical and the experimental spectra is more important. However, it must be noticed that these amplitudes are quite small compared to the ones of test #3, but the predicted order of magnitude is well verified.

It can be concluded that, if the mechanical and the electromagnetic characteristics of the materials that make the irradiated inhomogeneous rod are known or measured, the response of this rod can be predicted with a good accuracy. The method has been applied to an irradiated rod but it can also be applied to an irradiated multilayer plate.

IV. CONCLUSIONS

A model based on Maxwell's equations, a heat equation, and thermoviscoelasticity theory has been established to compute the particle velocity due to the acoustic field generated in a multilayer rod irradiated by pulsed microwaves and embedded in a rectangular waveguide or in a uniformly irradiated multilayer plate.

The theory has been validated experimentally by means of three different irradiated rod specimens used in four tests. The two first tests have been performed by irradiating two homogeneous rods made up of a PVC material or a rigid foam. The results of these two tests have been used to evalu-

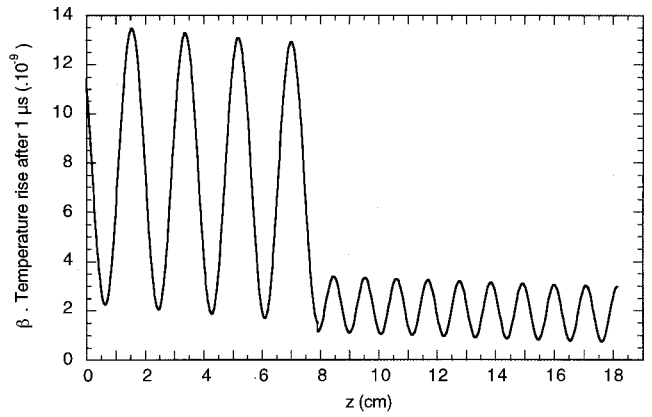


FIG. 9. Product of the maximum temperature rise after one pulse by the coefficient of linear thermal expansion β for the specimen foam/PVC (test #4).

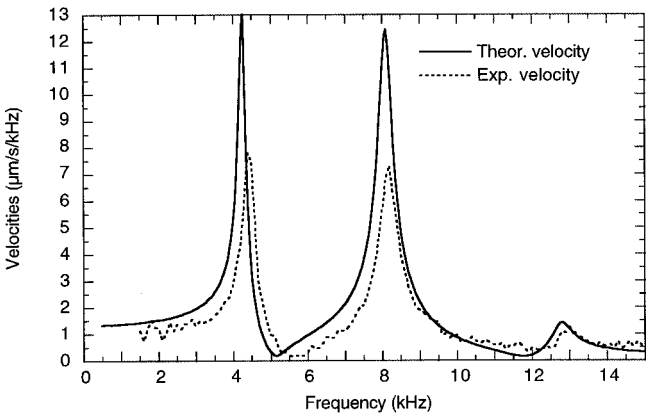


FIG. 11. Theoretical and experimental velocity spectra for the foam/PVC specimen (test #4).

ate and verify the mechanical and electromagnetic characteristics of these materials. Then, a third specimen has been made by gluing two rods made up of PVC and foam. Two tests have been performed on this specimen (by irradiating the PVC end or the foam end). The measured velocities have been compared to the predicted ones. The comparisons of the experimental and the theoretical spectra for each test show a good agreement. This fact validates the theory developed in the previous section.

ACKNOWLEDGMENTS

The authors are indebted to the *Conseil Régional d'Aquitaine* and the *Délégation Régionale à la Recherche et à la Technologie d'Aquitaine* for their financial support.

[1] D. A. Hutchins, ''Ultrasonic generation by pulsed laser,'' *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic Press, New York, 1988), Vol. XVIII.
[2] J. E. Michaels, ''Thermally induced elastic wave propagation in slender bars,'' *Proceedings of the 3rd US National Congress on Applied Mechanics* (American Society Mechanical Engineers, New York, 1958), p. 209.
[3] R. M. White, ''Generation of elastic waves by transient surface heating,'' J. Appl. Phys. **34**, 3559–3567 (1963).
[4] L. S. Gournay, ''Conversion of electromagnetic to acoustic energy by surface heating,'' J. Acoust. Soc. Am. **40**, 1322–1330 (1966).
[5] D. E. Borth and C. A. Cain, ''Theoretical analysis of acoustic signal generation in materials irradiated with microwave energy,'' IEEE Trans. Microwave Theory Tech. **MTT-25**, 944–953 (1977).
[6] L. M. Lyamshev and B. I. Chelnokov, ''Sound generation in a solid by penetrating radiation,'' Sov. Phys. Acoust. **29**, 220–225 (1983).
[7] B. Perrin, B. Bonello, J. C. Jeannet, and E. Romatet, ''Interferometric detection of hypersound waves in modulated structures,'' Prog. Nat. Sci. **6**, S444–S448 (1996).
[8] C. Rossignol, ''Etude théorique et numérique d'expériences d'acoustique picoseconde,'' Ph.D. thesis, Pierre & Marie Curie University, 2000 (in French) [English translation: Theoretical and numerical study of picosecond acoustic experiments].
[9] F. Enguehard and L. Bertrand, ''Effects of optical penetration and laser pulse duration on laser generated longitudinal acoustic waves,'' J. Appl. Phys. **82**, 1532–1538 (1997).
[10] B. Hosten and P. A. Bernard, ''Ultrasonic wave generation by time-gated microwaves,'' J. Acoust. Soc. Am. **104**, 860–866 (1998).
[11] C. Bacon, B. Hosten, and P. A. Bernard, ''Acoustic wave generation in viscoelastic rods by time-gated microwaves,'' J. Acoust. Soc. Am. **106**, 195–201 (1999).
[12] B. Hosten and C. Bacon, ''Measurement of complex young moduli of composite materials by time-gated microwaves,'' *Review of Progress in Quantitative Non Destructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 2000), Vol. 19, Part I, pp. 1113–1120.
[13] C. Bacon, B. Hosten, and E. Guilliorit, ''One-dimensional prediction of the acoustic waves generated in a multilayer viscoelastic body by microwave irradiation,'' J. Sound Vib. **238**, 853–867 (2000).
[14] C. Bacon, E. Guilliorit, B. Hosten, and D. Chimenti, ''Acoustic waves generated by pulsed microwaves in viscoelastic rods: modeling and experimental verification,'' J. Acoust. Soc. Am. **110**, 1398–1407 (2001).

Bulk reaction modeling of ducts with and without mean flow

S. K. Kakoty[a] and V. K. Roy[b]
Department of Mechanical Engineering, IIT Guwahati, North Guwahati, Guwahati-39, Assam, India

(Received 19 November 2001; revised 4 May 2002; accepted 7 May 2002)

A general formulation for analysis of sound field in a uniform flow duct lined with bulk-reacting sound-absorbing material is presented here. Presented theoretical model predicts the rate of attenuation for symmetric as well as asymmetric modes in rectangular duct lined with loosely bound (bulk-reacting) sound-absorbing material, which allows acoustic propagation through the lining. The nature of attenuation in rectangular ducts lined on two and four sides with and without mean flow is discussed. Computed results are compared with published theoretical and experimental results. The presented model can be used as guidelines for the acoustic design of silencers, air-conditioning ducts, industrial fans, and other similar applications. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1489433]

PACS numbers: 43.20.Hq, 43.20.Mv, 43.55.Rg, 43.20.Bi [RR]

LIST OF SYMBOLS

a_o	velocity of air at atmospheric temperature (m/s)
A_1, A_2, A_3, A_4	constants in Eq. (2)
b	half of duct width (m)
B_{1i}, B_{3i}, B_{4i}	constants in Eq. (11)
d	half of duct height (m)
D/Dt	$\partial/\partial t + U\partial/\partial z$ (total or substantive derivative)
∇^2	$\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$
f	frequency
h_o	propagation coefficient in lining
h_{xi}	propagation coefficient in ith lining in the x direction
h_{yi}	propagation coefficient in ith lining in the y direction
j	$\sqrt{-1}$
k_o	wave number in free air space
k_x	wave number in the x direction (normal to duct axis)
k_y	wave number in the y direction (normal to duct axis)
k_z	common wave propagation coefficient (parallel to duct axis)
l_x	lining thickness in the x direction (m)
l_y	lining thickness in the y direction (m)
m	trigonometric constant (integer) showing mode order in the x direction
M	mean flow Mach number
n	trigonometric constant (integer) showing mode order in the y direction.
p	acoustic pressure in central air space (Pa)
p_i	acoustic pressure in ith lining (Pa)
t	time (s)
U	mean flow velocity of fluid along duct axis (m/s)
u_{xi}, u_{yi}	acoustic velocity in ith lining in the x direction (m/s)
Y	characteristic impedance of the lining
Y_o	characteristic impedance of atmospheric air
ω	angular frequency (rad/s)
x,y,z	three-dimensional space coordinates
ρ_l	complex density of lining
ρ_o	density of atmospheric air (kg/m^3)
σ	specific flow resistance (N-s/m^4)
ϕ	nondimensional number

Subscripts

I	pertains to lining number: 1,2,3,4
x_i	pertains to the x direction in ith lining
y_i	pertains to y-direction in ith lining
o	pertains to value at atmospheric temperature
l	pertains to value for lining
x,y	pertains to transverse directions of duct
z	pertains to axial direction of duct

Superscripts

$+$	pertains to progressive wave
$-$	pertains to reflective wave

I. INTRODUCTION

There are two approaches for analysis of lined ducts

(i) Local reaction approach, where the lining is assumed to be locally reacting.[1]

(ii) Bulk reaction approach, which allows propagation of waves in the lining along the axis of the duct.[2]

Of these, the first one is more widely used for the simple reason of comparative ease of analysis. Local reaction modeling for rectangular ducts with and without mean flow is dealt with by many authors.[1,3–6] Bulk reaction modeling is available for circular ducts and rectangular ducts lined on

[a]Electronic mail: skkakoty@hotmail.com
[b]Scientist, Advanced Systems Laboratory, Hyderabad, India.

two opposite walls.[2,7–12] Very few authors have considered mean-flow effects for the bulk reaction modeling.[8] It is generally agreed that the lowest possible mode is not always the least attenuated.[8,9] Asymmetric modes are generally ignored, with a few exceptions.[5]

Kurze and Ver[7] considered nonisotropy in the lining material and developed equations which were later discussed by Wassilieff[9] and it was shown that the equations developed by Kurze and Ver[7] could be reduced to that of Scott's[2] equations developed for two-side lined ducts with only the assumption of isotropic property of the lining material. Higher-order modes were also discussed by Wassilieff[9] and a comparison was made between experimental and theoretical results. Christie[13] introduced solution schemes for bulk as well as the local reaction model. It was shown that using the Newton–Raphson scheme, it was possible to find the appropriate solution for the transcendental equations developed by Morse[1] and Scott.[2] Astley and Cummings[10] presented finite-element formulation for rectangular ducts with and without mean flow. A good agreement was found between their experimental and computed results. Bies *et al.*[11] presented closed-form formulations for rectangular and circular ducts with and without mean flow considering limp mass on the surface of the lining material, and demonstrated that when sound-absorbing material is covered with the limp mass, there is very little effect of direction of mean flow. Cummings and Sormaz[12] presented a theoretical model describing sound propagation in two-dimensional infinite planar silencers incorporating any number of bulk-reacting anisotropic sound-absorbing splitters with mean flow. A computational scheme was outlined to solve the equations developed by them. An analytical solution was presented for the wave propagation in two-dimensional rectangular-lined ducts with laminar mean flow by Gogate and Munjal.[14–16] Wu and Hamilton[17] described an analytical investigation of the propagation of sound in a lined circular duct with sheared mean ambient flow. Cummings[18] presented a review of research in the area of sound transmission through duct walls, especially that conducted over the past two decades. Finite-element predictions and corresponding measurements for sound attenuation and phase speed in a circular duct lined with poroelastic noise control foams were presented by Kang and Jung.[19]

It has been found that many of the works referred to here are based on finite-element formulation. However, these formulations are not very generalized in the sense that either all types of modes are not included or mean flow is not considered. In the present work, a generalized bulk reaction model has been developed for rectangular ducts lined on all four sides, considering asymmetric as well as symmetric modes, for a moving medium. The resulting set of equations reduces to the equation derived by others[2] in the past for different simpler cases. The Newton–Raphson scheme has been used for the simultaneous solution of the transcendental equations. The computed results of transmission loss, TL, are compared with the available theoretical as well as experimental observations. A brief study on the effects of higher-order mode, flow resistivity, and mean flow is presented. Results of two- and four-side lined ducts are also compared.

II. THEORY

A. Assumptions

There are several assumptions made in the theory which require some justification, and so these will be discussed first. The basic theoretical model consists of hard-walled rectangular duct, lined on two or four sides as the case may be. Lining thickness is uniform on the surface of the walls. Lining properties are considered to be isotropic. In the airway, uniform flow exists either in the same direction or in the direction opposite that of the sound propagation. To obtain a simple demonstration of the effects of the derived properties, the mean flow within the duct is taken to be uniform. Eversman[20] has shown that, for shear flow in circular ducts with locally reacting walls, the boundary layer thickness has little effect on attenuation for Mach numbers below 0.2. Assuming that bulk-reacting liners also behave in much the same way, one may allow the boundary layer to become infinitely thin for small Mach numbers, thus giving a uniform flow profile at the lining surface. In addition, the duct length is assumed to be infinite, so that there is no end reflection. And although the perforated ducts/silencers with various area discontinuities are in use nowadays, to simplify the derivation, only plain duct is considered here.

B. Propagation of sound through rectangular duct lined on all four sides with moving medium

In the theory that follows, the common path of acoustic wave is traced through composite passage of central air space and lining (see Fig. 1). There will be a common wave propagation constant and wave velocity for lining and central air space in the axial direction of duct and therefore, a common particle displacement at the air space and lining interface. The notations are given in the list of symbols.

(a) The convective wave equation in the free space is given by

$$\left(\frac{D^2}{Dt^2}-a_0^2\nabla^2\right)p=0. \tag{1}$$

The general solution to Eq. (1) is expected as

$$p=p(x,y,z,t)=(e^{-jk_xx}+A_3e^{jk_xx})(e^{-jk_yy}+A_4e^{jk_yy}) \times(A_1e^{-jk_z^+z}+A_2e^{jk_z^-z})e^{j\omega t}, \tag{2}$$

where k_x, k_y are the propagation constants along the x and y directions.

k_z^+, k_z^- are the propagation constants along the z direction for progressive (along the direction of mean flow) and reflective waves (in a direction opposite to mean flow), respectively.

From the compatibility condition

$$k_x^2+k_y^2+k_z^2=(k_o+Mk_z)^2,$$

or

$$k_z^{\pm}=\frac{\mp Mk_o+[k_o^2-(1-M^2)(k_x^2+k_y^2)]^{1/2}}{1-M^2}, \tag{3}$$

where $M=U/a_o$ = mean flow Mach number and $k_o=\omega/a_o$ = wave number (in free space).

For the reasoning given by Munjal[21] for the local reaction model, it can be shown that there will be a k_x^+ and k_y^+ corresponding to k_z^+, and k_x^- and k_y^- corresponding to k_z^-. So, forward- and backward-moving waves are considered separately to find k_x^+, k_y^+, k_z^+, and k_x^-, k_y^-, k_z^-. Equation (3) is therefore rewritten as

$$k_z^{\pm}=\frac{\mp Mk_o+[k_o^2-(1-M^2)(k_x^{\pm 2}+k_y^{\pm 2})]^{1/2}}{1-M^2}. \tag{4}$$

Thus, pressure for forward-moving waves will be

$$p^+=A_1(e^{-jk_x^+x}+A_3^+e^{jk_x^+x}) \times(e^{-jk_y^+y}+A_4^+e^{jk_y^+y})e^{-jk_z^+z}e^{j\omega t}, \tag{5}$$

and pressure for backward-moving waves will be

$$p^-=A_2(e^{-jk_x^-x}+A_3^-e^{jk_x^-x}) \times(e^{-jk_y^-y}+A_4^-e^{jk_y^-y})e^{jk_z^-z}e^{j\omega t}. \tag{6}$$

Components of momentum equation are

$$\rho_0\frac{Du_x^{\pm}}{Dt}+\frac{\partial p^{\pm}}{\partial x}=0, \tag{7a}$$

$$\rho_0\frac{Du_y^{\pm}}{Dt}+\frac{\partial p^{\pm}}{\partial y}=0, \tag{7b}$$

$$\rho_0\frac{Du_z^{\pm}}{Dt}+\frac{\partial p^{\pm}}{\partial z}=0. \tag{7c}$$

Particle velocity components u_x, u_y, u_z can be written in the form of Eq. (5). Substituting the pressure from Eq. (5) in the above three components of momentum equation and equating the coefficients of the different exponentials separately to zero, particle velocity in the y direction is found to be

$$u_y^+=\frac{1}{\rho_o}\frac{k_y^+}{a_o(1-Mk_z^+/k_o)}A_1(e^{-jk_x^+x}+A_3^+e^{jk_x^+x}) \times(e^{-jk_y^+y}-A_4^+e^{jk_y^+y})e^{-jk_z^+z}e^{j\omega t} \tag{8}$$

and similar expressions for u_x^+ and u_z^+.

(b) Wave equation in lining.

With the application of conservation of mass and dynamical equilibrium inside the lining, the wave equation in the ith section of the lining is given by

$$(\nabla^2+h_o^2)p_i=0, \quad i=1,2,3,4, \tag{9}$$

where h_o is a complex quantity and related to the attenuation and phase velocity of sound wave. This quantity is expressed as

$$h_o=\beta_o-j\alpha_o,$$

where α_o is the attenuation constant and β_o is the wavelength constant of an acoustic wave in the lining. Propagation constants along the duct axis (k_z) will be the same for both linings and free central air space. So,

$$h_o^2=h_{xi}^{\pm 2}+h_{yi}^{\pm 2}+k_z^{\pm 2}, \quad i=1,2,3,4. \tag{10}$$

Acoustic pressure for the forward-moving part of the wave in the ith lining will be

$$p_i^+=B_{1i}(e^{-jh_{xi}^+x}+B_{3i}^+e^{jh_{xi}^+x}) \times(e^{-jh_{yi}^+y}+B_{4i}e^{jh_{yi}^+y})e^{-jk_z^+z}e^{j\omega t}. \tag{11}$$

Corresponding acoustic velocity in the y direction will be

$$u_{yi}^+=-\frac{j}{\rho_l\omega}\frac{\partial p_i}{\partial y}=\frac{h_{yi}^+B_{1i}}{\rho_l\omega}(e^{-jh_{xi}^+x}+B_{3i}^+e^{jh_{xi}^+x}) \times(e^{-jh_{yi}^+y}-B_{4i}e^{jh_{yi}^+y})e^{-jk_z^+z}e^{j\omega t}. \tag{12}$$

Similarly, u_{xi}^+ and u_{zi}^+ can be found out.

(c) Boundary conditions.

The first boundary condition is that fluid should possess zero normal particle velocity at the outer wall. Considering lining number 1 ($i=1$) and the y direction for progressive wave, i.e.,

$$\text{at } y=d+l_y, \quad u_{y1}^+=0.$$

So, from Eq. (12), $e^{-jh_{y1}^+(d+l_y)}=B_{41}^+e^{-jh_{y1}^+(d+l_y)}$, which gives

$$B_{41}^+=e^{-2jh_{y1}^+(d+l_y)}. \tag{13}$$

Let the common particle displacement at free space and lining interface be denoted as η. So, the particle velocity in lining number 1 in the y direction for progressive wave is given by

$$u_{y1}^+=\frac{\partial\eta}{\partial t},$$

and particle velocity in central air space in the y direction for progressive wave is given by

$$u_y^+=\frac{D\eta}{Dt}.$$

Eliminating η, one can have

$$\frac{Du_{y1}}{Dt}=\frac{\partial u_y^+}{\partial t},$$

which gives

$$u_y^+=(1-Mk_z^+/k_o)u_{y1}^+. \tag{14}$$

The second boundary condition is that the acoustic pressure at the interface of the central part of the duct and lining (number 1) should be continuous, i.e.,

$$\text{at } y=d, \quad p^+=p_1^+.$$

So,

$$\frac{u_y^+}{p^+}=(1-Mk_z^+/k_o)\frac{u_{y1}^+}{p_1^+}. \tag{15}$$

Substituting p^+, u_y^+, p_1^+, and u_{y1}^+ from Eqs. (5), (8), (11), and (12) into Eq. (15)

$$\frac{k_y^+(-e^{-jk_y^+d}+A_4e^{jk_y^+d})}{\rho_o(e^{-jk_y^+d}+A_4^+e^{jk_y^+d})}$$

$$=-(1-Mk_z^+/k_o)^2\frac{h_{y1}^+(e^{-jh_{y1}^+d}-B_{41}e^{jh_{y1}^+d})}{\rho_l(e^{-jh_{y1}^+d}+B_{41}e^{jh_{y1}^+d})}. \quad (16)$$

Substituting B_{41} from Eq. (13)

$$\frac{k_y^+(-e^{-jk_y^+d}+A_4e^{jk_y^+d})}{\rho_0(e^{-jk_y^+d}+A_4^+e^{jk_y^+d})}$$

$$=-j(1-Mk_z^+/k_o)^2\frac{h_{y1}^+}{\rho_l}\tan(h_{y1}^+l_y). \quad (17)$$

At the outer wall (say adjacent to lining number 1), the particle velocity in the x direction will be zero, i.e.,

$$\text{at } x=b+l_x \text{ and } x=-(b+l_x), \quad u_{x1}^+=0,$$

or,

$$e^{-jh_{x1}^+(b+l_x)}-B_{31}^+e^{jh_{x1}^+(b+l_x)}$$

$$=0=e^{jh_{x1}^+(b+l_x)}-B_{31}^+e^{jh_{x1}^+(b+l_x)},$$

which yields $B_{31}=-1$ and $e^{-jh_{x1}^+(b+l_x)}+e^{jh_{x1}^+(b+l_x)}=0$. So,

$$\cos(h_{x1}^+(b+l_x))=0.$$

Hence,

$$h_{x1}^+=\frac{(2m+1)\pi}{2(b+l_x)}, \quad m=0,1,2,3,4,\ldots. \quad (18)$$

And similarly, for lining number 3

$$\text{at } x=(b+l_x) \text{ and } x=-(b+l_x), \quad u_{x3}^+=0.$$

So, one can arrive at

$$B_{33}^+=-1 \text{ and } e^{-jh_{x3}^{+(b+l_x)}}+e^{jh_{x3}^{+(b+l_x)}}=0,$$

and thereby

$$h_{x3}^+=\frac{(2m+1)\pi}{2(b+l_x)}, \quad m=0,1,2,3,4,\ldots. \quad (19)$$

As h_{x1}^+ and h_{x3}^+ are the same, so a third notation h_{x1-3}^+ (which indicates either h_{x1} or h_{x3}) has been used so that both notations can be used interchangeably and either value can be used in computation.

So, from Eqs. (18) and (19)

$$h_{x1}^+=h_{x3}^+=h_{x1-3}^+=\frac{(2m+1)\pi}{2(b+l_x)}, \quad (20)$$

and similarly

$$h_{y1}^+=h_{y3}^+=h_{y1-3}^+.$$

Now, one can rewrite Eq. (10) in the following form:

$$h_o^2=h_{x1-3}^{+2}+h_{y1-3}^{+2}+k_z^{+2}. \quad (21)$$

At the interface of lining number 3 and central air space, particle velocity and acoustic pressure will be continuous, i.e.,

$$\text{at } y=-d \quad u_{y3}^+=u_y^+ \text{ and } p_3^+=p^+,$$

from which one gets

$$\frac{k_y^+(e^{jk_y^+d}-A_4^+e^{-jk_y^+d})}{\rho_o(e^{jk_y^+d}+A_4^+e^{-jk_y^+d})}$$

$$=-j(1-Mk_z^+/k_o)^2\frac{h_{y1-3}^+}{\rho_l}\tan(h_{y1-3}^+l_y). \quad (22)$$

From Eqs. (17), (20), and (22)

$$A_4^{+2}=1 \quad \text{or} \quad A_4^+=\pm 1.$$

When $A_4^+=+1$, either of Eqs. (17) and (22) becomes

$$\frac{k_y^+}{\rho_o}\tan(k_y^+d)=-(1-Mk_z^+/k_o)^2\frac{h_{y1-3}^+}{\rho_l}\tan(h_{y1-3}^+l_y). \quad (23)$$

When $A_4^+=-1$, either of Eqs. (17) and (22) becomes

$$\frac{k_y^+}{\rho_o}\cot(k_y^+d)=(1-Mk_z^+/k_o)^2\frac{h_{y1-3}^+}{\rho_l}\tan(h_{y1-3}^+l_y). \quad (24)$$

Similarly, one can derive

$$h_o^2=h_{x2-4}^{+2}+h_{y2-4}^{+2}+k_z^{+2}, \quad (25)$$

where $h_{y2-4}^+=[(2n+1)\pi]/[2(d+l_y)]$, $n=0,1,2,3,4...$,

$$\frac{k_x^+}{\rho_o}\tan(k_x^+b)=-(1-Mk_z^+/k_o)^2\frac{h_{x2-4}^+}{\rho_l}\tan(h_{x2-4}^+l_x), \quad (26)$$

$$\frac{k_x^+}{\rho_o}\cot(k_x^+b)=(1-Mk_z^+/k_o)^2\frac{h_{x2-4}^+}{\rho_l}\tan(h_{x2-4}^+l_x). \quad (27)$$

Equations (23) and (26) correspond to symmetric modes and Eqs. (24) and (27) correspond to asymmetric modes. Equations (23) and (26) or Eqs. (24) and (27), along with Eqs. (4), (21), and (25), can be solved for k_x^+, k_y^+, k_z^+, h_{y1-3}^+, and h_{x2-4}^+. All these equations derived are for forward-moving waves. Similar expressions can be derived for k_x^-, k_y^-, k_z^-, h_{y1-3}^-, and h_{x2-4}^- for reflective waves. However, these may be readily obtained from equations derived for forward-moving waves by substituting the superscript $(+)$ by $(-)$ and M by $-M$. Putting $M=0$ and removing the superscripts $(+)$ or $(-)$, one can arrive at the equations for zero mean flow.

As the metallic wall of duct has very high (infinite) impedance, if there is no lining along any direction (x or y) then the contribution of the wave number in that direction can be reduced to zero. Therefore, equations derived in this section can be reduced to the case of ducts lined on two opposite sides with and without mean flow. Substituting zero for k_x and l_x, Eqs. (23), (24), (26), and (27) reduce to two simpler equations which, without mean flow, further reduce to

$$\frac{k_y}{\rho_o}\tan(k_yd)=-\frac{h_y}{\rho_l}\tan(h_y,l_y), \quad (28)$$

$$\frac{k_y}{\rho_o}\cot(k_yd)=\frac{h_y}{\rho_l}\tan(h_yl_y). \quad (29)$$

Equation (28) is simply what Scott[2] developed in his analysis. This equation applies to symmetric modes only. Equation (29) applies to asymmetric case. The left-hand sides of the equations are the same for both the local and bulk reaction model. But, the right-hand sides are different, as propagation of sound in the axial direction through the linings in the local reaction model is not considered.

III. SOLUTION SCHEME

Various efforts have been made to solve the equations developed for the bulk reaction model. One of that is the finite-element method used by Astley and Cummings.[10] The first step in numerical computation with the finite-element method is the consideration of a greater number of appropriate elements. Since the size of the overall matrix is proportional to the number of nodes, considerably more computation time and more core memory are required. It has been found that if the number of independent variables is N, then the core memory requirement is proportional to N^2 and solution time is proportional to N^3. To reduce the size of resulting stiffness matrix of ducts, Craggs[22] solved this type of problem by a combination of transfer matrix and matrix condensation method with the finite-element method. With the help of this methodology, complex geometry problems with numerous junctions could be solved. But, numerical stability was not ensured here in case of singular stiffness matrix.

Here, the Newton–Raphson scheme has been used to solve the equations developed. The method requires a very close approximation of starting value for iteration. To get the appropriate starting value, the method outlined by Christie[13] has been adopted. According to Christie,[13] suitable approximation to start the iteration at low frequencies can be obtained by replacing the tangents by their arguments ($\tan x \cong x$) in the equations developed for the (0,0) mode.

To start with, only symmetric modes are considered. So, Eqs. (23) and (26) will be

$$\frac{k_y^2 d}{\rho_o} = -\frac{h_{y1-3}^2 l_y}{\rho_l}, \tag{30}$$

$$\frac{k_x^2 b}{\rho_0} = -\frac{h_{x2-4}^2 l_x}{\rho_l}. \tag{31}$$

Equations (30) and (21) give

$$\frac{k_y^2 d}{\rho_o} = -\frac{1}{\rho_l}\left[h_o^2 - \left(\frac{(2m+1)\pi}{2(b+l_x)}\right)^2 - k_z^2\right] l_y. \tag{32}$$

Equations (31) and (25) give

$$\frac{k_x^2 b}{\rho_o} = -\frac{1}{\rho_l}\left[h_o^2 - \left(\frac{(2n+1)\pi}{2(d+l_y)}\right)^2 - k_z^2\right] l_x. \tag{33}$$

Substituting k_x and k_y from Eqs. (32) and (33) in Eq. (3) with $M=0$, one gets the expression for k_z as

$$k_z^2 = \frac{(k_o^2 + R_1 h_o^2 - R_2)}{(1+R_1)}, \tag{34}$$

where

$$R_1 = \frac{\rho_o l_x}{\rho_l b} + \frac{\rho_o l_y}{\rho_l d},$$

$$R_2 = \frac{\rho_o l_x}{\rho_l b}\left(\frac{(2n+1)\pi}{2(d+l_y)}\right)^2 + \frac{\rho_o l_y}{\rho_l d}\left(\frac{(2m+1)\pi}{2(b+l_x)}\right)^2.$$

The scheme of computation is described here

(1) Making use of physical properties of lining material, the complex propagation coefficient (h_o), characteristic impedance (Y), and complex density (ρ_l) of the sound-absorbing material are calculated. Delany and Bazely[23] have shown that for fibrous materials, empirical expressions may be used to calculate propagation coefficient, characteristic impedance, and complex density in terms of a nondimensional parameter $\phi(=\omega\rho_o/2\pi\sigma)$ as below.

$$\frac{Y}{Y_o} = (1 + c_1\phi^{c_2}) - j(c_3\phi^{c_4}),$$

where $Y_o = \rho_o a_o$,

$$\frac{h_o}{k_o} = c_5\phi^{c_6} + j(1 + c_7\phi^{c_8}), \quad \rho_l = \frac{Y h_o}{\omega}.$$

Constants c_1–c_8 are the best-fit values of the Delany and Bazely model for a given lining material. Here, in all the calculations, except for Fig. 4, polyurethane foam (used by Wassilieff[9] and Christie[13]) is used for which c constants of the Delany and Bazely model are $c_1=0.0571$, $c_2=-0.754$, $c_3=0.087$, $c_4=-0.732$, $c_5=0.189$, $c_6=-0.595$, $c_7=0.0978$, $c_8=-0.70$. In Fig. 4, calculation has been done for pink foam, for which the c constants are $c_1=0.302$, $c_2=-0.322$, $c_3=0.10$, $c_4=-0.631$, $c_5=0.193$, $c_6=-0.505$, $c_7=0.275$, $c_8=-0.359$.

(2) Calculate the first approximation values of k_x, k_y, and k_z from Eqs. (32), (33), and (34), keeping $M=0$.

(3) Refine the first approximation values using the Newton–Raphson method. Converged value for a particular frequency is treated as first approximation value for the next higher frequency.

(4) The first approximation values of k_x, k_y, and k_z for the

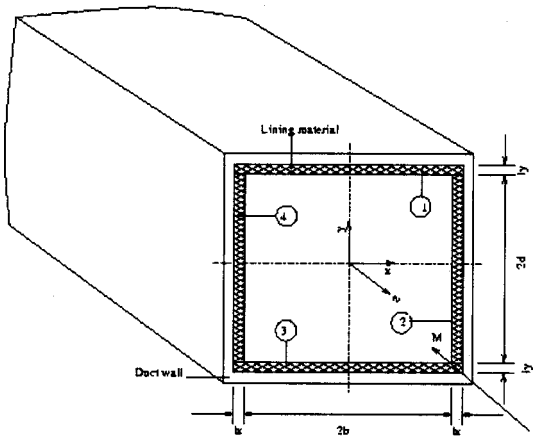


FIG. 1. Four-side-lined duct with mean flow.

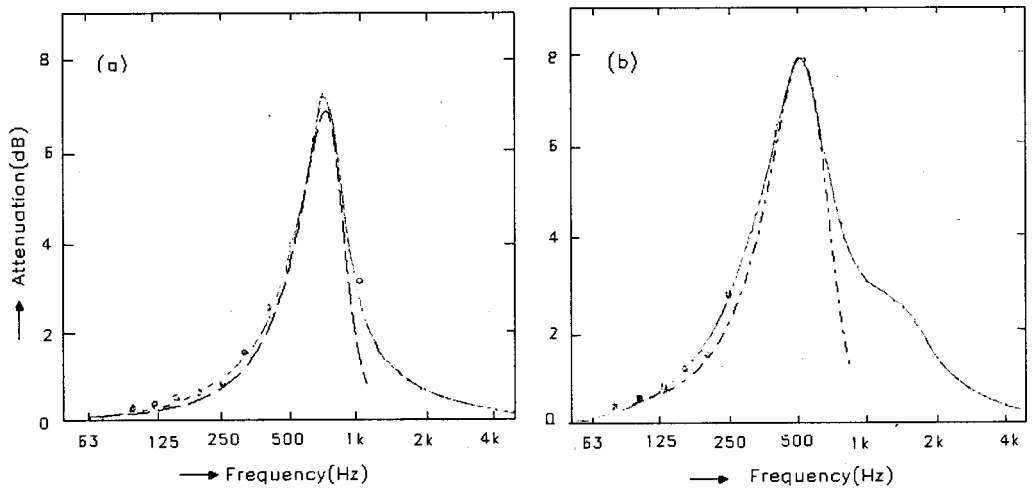


FIG. 2. Attenuation versus frequency for polyurethane foam for lining thickness and duct width (a) 50 mm, 500 mm,..., measured result by Wassilieff, ———, theoretical result by Wassilieff (Ref. 9), ---, calculated result from present model (b) 100 mm, 400 mm,..., measured result by Wassilieff, ———, theoretical result by Wassilieff (Ref. 9), -·-·-, calculated result from present model. Specific flow resistivity for both cases is 8500 Pa-s/m^2.

case with mean flow are taken from the converged values of the case without mean flow for corresponding frequency.

(5) For waves moving in the opposite direction, the value of M is reversed in all the equations, which are then solved similarly.

(6) Absolute attenuation is found out as TL$=8.69*$ Im$(k_z)*$ duct width (or per unit length in meter).

IV. RESULTS AND DISCUSSION

Attenuation is the imaginary part of the common propagation coefficient k_z as per the definition. Using the methodology outlined above, attenuation is calculated for the desired case. For two-side-lined ducts without mean flow, equations are solved simultaneously by the Newton–Raphson scheme with an accuracy of 10^{-5}. It has been observed that, except at some low frequencies, it took barely three to four iterations to converge for this level of accuracy. This shows that the initial approximation is quite satisfactory. At sufficiently low frequencies, the number of iterations required is found to be about 10 to 15. The attenuation plot is obtained by plotting TL with respect to frequency. Results are presented in Figs. 2–7 illustrating the nature of attenuation.

A. Comparison with other works

In this work, no experimental setup is used to validate the present theoretical model. So, first of all, effort has been made to compare the present work with some published results. In Figs. 2(a) and (b), plots are shown side by side for two different dimensions of duct and lining thickness used by Wassilieff.[9] Good agreement has been observed in both cases. Figure 3(a) is the theoretical result presented by Christie[13] for three different cases. Figure 3(b) is the theoretical result obtained with the present mathematical model. In all three cases of Fig. 3, average error with respect to Christie's finding was around (5%–6%). It was found that at low flow resistivity (cases I and II), the result of the present theoretical model is not much different from that of Christie,

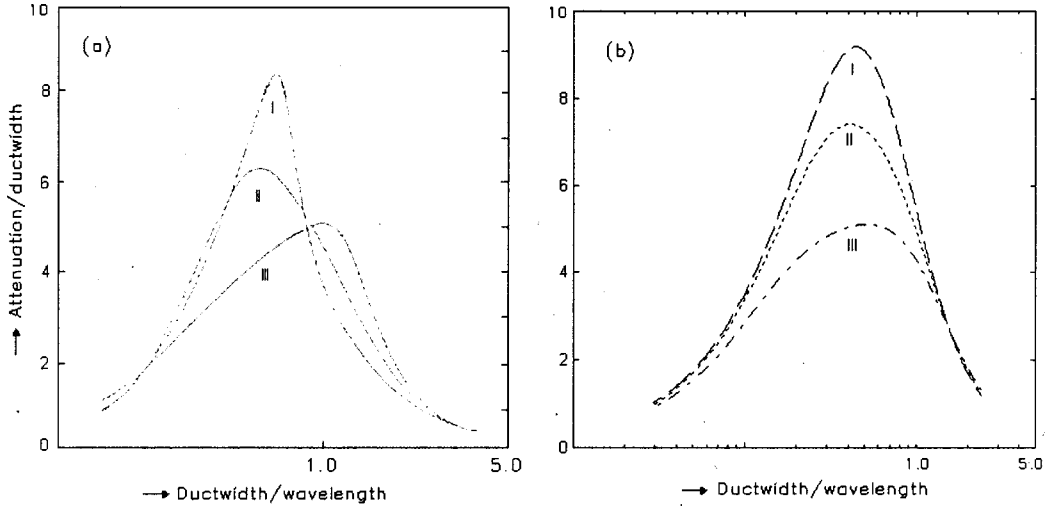


FIG. 3. Attenuation versus frequency for two-side-lined duct for polyurethane foam for duct dimension $2d=0.5334$ m, $2b=0.5334$ m, and lining thickness $l_y=0.1524$ m. Specific flow resistivity (Pa-s/m^2) for (I) 4000; (II) 8000; (III) 16 000.

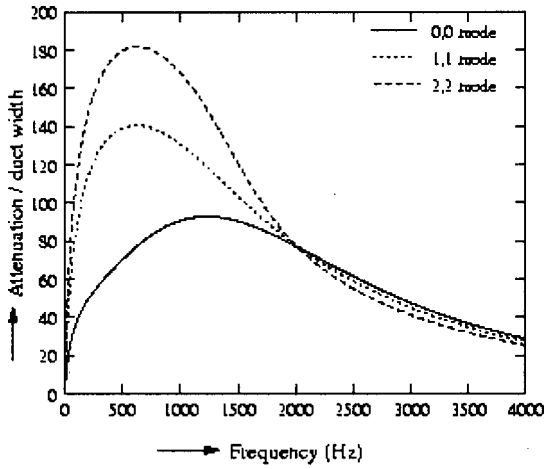


FIG. 4. Attenuation versus frequency for duct size $l_x=0.0495$ m, $l_y=0.0495$ m, $d=0.0755$ m, $b=0.0505$ m. Specific flow resistivity 9050 Pa-s/m^2. (1) 0,0, (2) 1,1, (3) 2,2 mode.

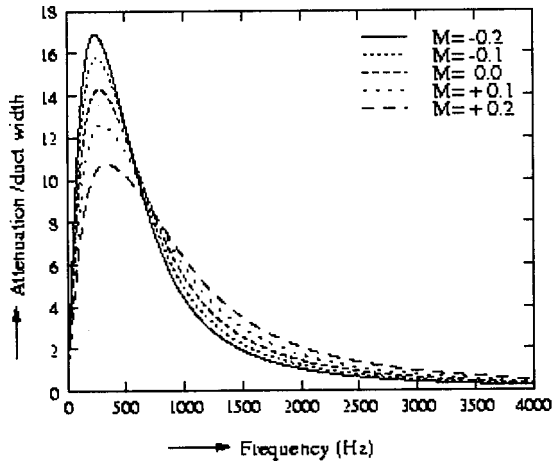


FIG. 6. Attenuation versus frequency for (0,0) mode for four-side-lined duct with cross section $2b=2d=0.5334$ m, $l_y=l_x=0.1524$ m, specific flow resistivity 10 000 Pa-s/m^2 with mean flow Mach number (1) $M=-0.2$. (2) $M=-0.1$. (3) $M=0.0$. (4) $M=+0.1$. (5) $M=+0.2$.

but at higher flow resistivity (case III), shift of peak attenuation towards the right is not found with the present model. However, in the case of two-side-lined ducts, attenuation value was found to shoot off (or else the plot is becoming erratic) in a higher frequency range, which could not be shown for the sake of clarity of existing plots. For this reason, plots are shown to be truncated in a higher frequency range.

B. Effect of higher modes

Many of the theoretical and experimental results have been published without any attempt to excite modes of preference. It has usually been assumed that the fundamental mode is both least attenuated and most strongly excited. So, most of the literature available is limited to fundamental modes only. This condition is reasonably met in air-conditioning and ventilation ducts, because small cross section allows propagation of plane waves only. However, there are situations in practice where conditions for plane wave may not be met. An important one is the acoustically lined duct of a quieted high-bypass turbojet engine nacelle, in which transverse dimensions of the duct permit propagation of many higher-order modes. In Fig. 4, three plots are shown. It has been observed that the fundamental mode is the least attenuated in the low-frequency region only. At higher frequencies, trends were different and it was found that not only can higher-order modes be strongly excited but also fundamental mode need not necessarily be least attenuated. Attenuation obtained for the (0,0) mode is in very good agreement with that obtained by Astley and Cummings.[10] Attenuation for two higher-order modes is also shown in this figure. Plots show that at low frequencies, the (0,0) mode is always least attenuated. But, at higher frequencies, the trend is opposite and (0,0) mode becomes highly attenuated after a crossover point and higher-order modes become least attenuated. It was also found that for the zero-order mode, maximum attenuation exists near about 1500 Hz, which gives a wavelength nearly four times that of the lining thickness

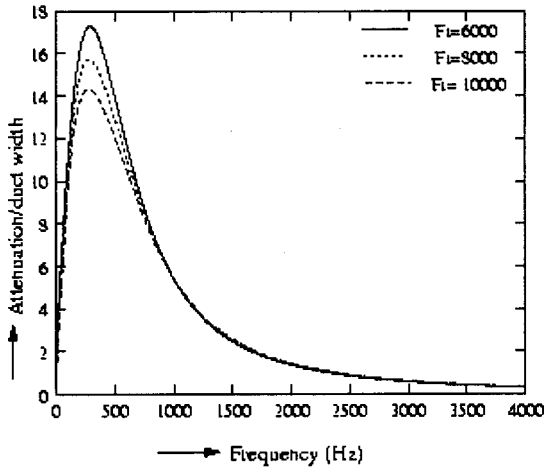


FIG. 5. Attenuation versus frequency for (0,0) mode for four-side-lined duct without mean flow with duct size $l_y=l_x=0.1524$ m, $2b=2d=0.5334$ m for different values of specific flow resistivity.

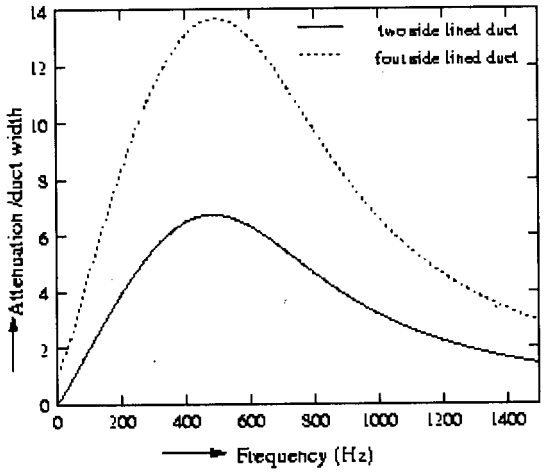


FIG. 7. Plot illustrating the nature of attenuation for two-side-lined duct and four-side-lined duct. For four-side-lined duct $l_y=l_x=0.05$ m, $2b=0.50$ m, $2d=0.50$ m. For two-side-lined duct $l_y=0.05$ m, $2b=0.50$ m, $2d=0.50$ m.

used in the duct. This is in agreement with experimental findings of King.[24,25]

C. Effect of flow resistivity

If the flow resistivity of the absorbent is changed, usually the attenuation increases at some frequencies and decreases at some other frequencies. So, there is no single value dependency of flow resistivity for maximum attenuation. It is, however, important to investigate the dependency of flow resistivity while other parameters are kept constant. Figure 5 shows the curves for attenuation for the (0,0) mode. Three curves are shown here for three different values of flow resistivity. In the low-frequency region, the effect of flow resistivity is well reflected, but in high-frequency region it is less pronounced. This theoretical finding is in agreement with several published results like Cummings[8] and Astley and Cummings.[10]

D. Effect of mean flow

Figure 6 shows that attenuation rate depends greatly on mean flow. Here, the Mach number is in the range of -0.2 to $+0.2$ for many reasons. First, it is unusual for silencers and mufflers to exceed this limit, as higher Mach numbers will cause erosion inside the ducts. Second, the uniform mean-flow approximation ceases to valid at higher Mach numbers. Here, positive mean flow corresponds to downstream propagation and negative Mach number corresponds to upstream propagation. In Fig. 6, five curves are shown. The main feature of interest is that downstream-propagating waves are progressively less attenuated at low frequencies. As the Mach number increases, attenuation further comes down before a crossover point where this trend reverses. Negative Mach number shows the opposite trend. Increasing Mach number negatively at low frequency, attenuation increases, but then decreases after the crossover point. There are several possible explanations for this trend. One possible explanation is that the uniform flow convects the sound wave, but the boundary layer refracts it as it passes into the porous material. The wave traveling downstream, because its acoustic phase velocity is increased, has less chance to be absorbed than that traveling upstream because the sound wave has traveled further due to increased relative velocity compared to zero mean flow. Mathematically the credit goes to the $D/Dt=\partial/\partial t+U(\partial/\partial z)$ term. At higher frequency, the refraction of the sound wave by the boundary layer into the lining starts to dominate over the convective effect and eventually this brings about an improvement in attenuation. The opposite effect occurs for upstream propagation.

E. Two- and four-side-lined duct

Figure 7 shows two plots for two- and four-side-lined ducts for attenuation without the effect of mean flow. The attenuation of the four-side-lined duct is found to be almost double that of the two-side-lined duct for the entire frequency range. Therefore, peak attenuation of the four-side-lined duct without mean flow can be estimated knowing the peak attenuation of the two-side-lined duct. However, the effect of mean flow has to be ascertained in this case prior to generalizing the trend.

V. CONCLUSION

A simple but accurate model is presented here for estimation of duct attenuation. It has been shown that the bulk reaction modeling of lined duct can be handled very easily while incorporating the wave propagation through lining material, too. The solution scheme is very simple, and the present results are in good agreement with published results, even with different solution schemes. Effects of flow resistivity, mean flow, and higher-order modes are studied in the light of the present sets of results. The study can be extended for other geometry, especially for circular ducts and rectangular ducts with an air gap between lining and rigid wall.

[1] P. M. Morse, ''The transmission of sound inside pipes,'' J. Acoust. Soc. Am. **11**, 205–210 (1939).
[2] R. A. Scott, ''The propagation of sound between walls of porous material,'' Proc. Phys. Soc. London **58**, 358–368 (1945).
[3] M. L. Munjal and U. S. Shirahatti, ''Analysis of lined ducts with mean flow, the application to dissipative mufflers,'' Trans. ASME, J. Vib., Acoust., Stress, Reliab. Des. **109**, 366–371 (1987).
[4] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[5] J. E. Kaiser, B. S. Shaker, and A. H. Nayfeh, ''Influence of liner thickness on wave propagation in ducts,'' J. Sound Vib. **37**(2), 169–183 (1974).
[6] A. Craggs, ''A finite element model for acoustically lined small rooms,'' J. Sound Vib. **108**(2), 327–337 (1986).
[7] U. J. Kurze and I. L. Ver, ''Sound attenuation in ducts with non-isotropic material,'' J. Sound Vib. **24**(2), 177–187 (1972).
[8] A. Cummings, ''Sound attenuation in ducts lined on two opposite walls with porous material, with some applications to splitter,'' J. Sound Vib. **49**(1), 9–35 (1976).
[9] C. Wassilieff, ''Experimental verification of duct attenuation modes with bulk reacting linings,'' J. Sound Vib. **114**(2), 239–251 (1987).
[10] R. J. Astley and A. Cummings, ''A finite element scheme for attenuation in ducts lined with porous material, comparison with experiment,'' J. Sound Vib. **116**(2), 239–261 (1987).
[11] D. A. Bies, C. H. Hansen, and G. E. Bridges, ''Sound attenuation in rectangular and circular cross-section ducts with flow and bulk-reacting liners,'' J. Sound Vib. **146**(1), 47–80 (1991).
[12] A. Cummings and N. Sormaz, ''Acoustic attenuation in dissipative splitter silencers containing mean flow,'' J. Sound Vib. **168**(2), 209–227 (1993).
[13] D. R. A. Christie, ''Theoretical attenuation of sound in a lined duct,'' J. Sound Vib. **17**(2), 283–286 (1971).
[14] G. R. Gogate and M. L. Munjal, ''Analytical solution of the laminar mean flow wave equation in a lined or unlined two-dimensional rectangular duct,'' J. Acoust. Soc. Am. **92**, 2915–2923 (1992).
[15] G. R. Gogate and M. L. Munjal, ''Sound propagation in ducts with bulk reacting lining in the presence of laminar mean flow,'' J. Acoust. Soc. Am. **99**, 1779–1782 (1996).
[16] G. R. Gogate and M. L. Munjal, ''Analytical solution of sound propagation in lined or unlined circular ducts with laminar mean flow,'' J. Sound Vib. **160**(3), 465–484 (1993).
[17] J. Wu and M. F. Hamilt, ''Propagation of sound in a lined circular duct with sheared mean flow,'' J. Acoust. Soc. Am. **96**, 3255 (1994).
[18] A. Cummings, ''Sound transmission through duct walls,'' J. Sound Vib. **239**(4), 731–765 (2001).
[19] Y. J. Kang and I. H. Jung, ''Sound propagation in circular ducts lined with noise control foams,'' J. Sound Vib. **239**(2), 255–273 (2001).
[20] W. Eversman, ''Effect of boundary layer on the transmission and attenuation of sound in an acoustically treated circular duct,'' J. Acoust. Soc. Am. **49**, 1372–1380 (1971).
[21] M. L. Munjal, *Acoustics of Ducts and Mufflers with Application to Exhaust and Ventilation System Design* (Wiley, New York, 1987).
[22] A. Craggs, ''The application of the transfer matrix method and matrix

condensation methods with finite elements to duct acoustics,'' J. Sound Vib. **132**(2), 393–402 (1989).
[23] M. E. Delany and E. N. Bazely, ''Acoustical properties of fibrous acoustical materials,'' Appl. Acoust. **3**, 105–116 (1970).
[24] A. J. King, ''Attenuation of sound in lined air ducts,'' J. Acoust. Soc. Am. **30**, 505–507 (1958).
[25] A. Bokor, ''Attenuation of sound in lined ducts,'' J. Sound Vib. **10**(3), 390–403 (1969).

Design of optimal configuration for generating A_0 Lamb mode in a composite plate using piezoceramic transducers

Sébastien Grondel[a)]
IEMN, UMR CNRS 9929, Département OAE, Université de Valenciennes et du Hainaut Cambresis, Le Mont Houy, 59313 Valenciennes Cedex 9, France

Christophe Paget
Aeronautics Division, FOI, Swedish Defence Research Agency, 172 90 Stockholm, Sweden

Christophe Delebarre and Jamal Assaad
IEMN, UMR CNRS 9929, Département OAE, Université de Valenciennes et du Hainaut Cambresis, Le Mont Houy, 59313 Valenciennes Cedex 9, France

Klas Levin
Aeronautics Division, FOI, Swedish Defence Research Agency, 172 90 Stockholm, Sweden

(Received 11 December 2001; revised 27 March 2002; accepted 4 April 2002)

This work concerned a technique for a health monitoring system based on the generation and sensing of Lamb waves in composite structures by thin surface-bonded piezoceramic transducers. The objective was to develop transducers that are adapted for the damage detection in orthotropic composites. The key problem with the investigated Lamb waves was to select a mode to be sensitive to the damage. A hybrid modeling technique was therefore used to conceive transducers that were adapted to achieve such a feature. This modeling technique enabled studying the influence of the transducer characteristics on the Lamb waves propagating in orthotropic plates. It was demonstrated that a Lamb mode could be generated dominantly to other modes by using a multi-element transducer. The effectiveness of this technique was successfully verified experimentally on composite plates. It was shown that the dominant Lamb mode, obtained by use of dual-element transducers, was an appropriate mode for successfully detecting a damage in composites. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1481062]

PACS numbers: 43.20.Ks, 43.20.Mv, 43.40.Le [JGH]

I. INTRODUCTION

With the increasing use of advanced composite materials in aerospace structures, it is of growing interest to develop built-in structural health-monitoring systems to ensure the integrity of these materials. One technique, which has a great potential for a quick and continuous inspection of aircraft structures, is based on piezoceramic transducers for generating and receiving Lamb waves. Recent studies performed on aluminum riveted plates,[1] wingbox composite structures,[2,3] and radomes[4] have proven the feasibility of this technique to detect various types of defects, such as fatigue cracks, or low-velocity impact damage. In these studies, thin piezoceramic transducers that generated Lamb waves were bonded onto the structure. In comparison with the conventional and wedged transducers, the thin piezoceramic transducer has significantly greater potential to be adapted for built-in structural health monitoring. Because of their small thickness these transducers can also be embedded[5] in composites. Although this research had already proved the immediate benefit of this technique, some additional issues had yet to be solved. In particular, the transducer design has to be improved for enhancing the detectability of the damage. This can be done by improving the generation and reception by transducers of those Lamb modes, which are sensitive to the damage.

In the last decades, a wide range of theoretical work has been done to better understand the propagation of Lamb waves by piezoceramic transducers. In most of these studies, the transducer excitation on the structure was known. Thus the problem consisted of solving the wave propagation equations with respect to the boundary conditions. The problem was solved by applying an integral transform technique where the result was transformed to the wavenumber or frequency domains. Viktorov[6] used the Fourier transform to deal with harmonic excitations by a wedge transducer on isotropic materials. Miklowitz[7] extended this method by using the Laplace transform to solve transient loading. This technique was also extended to orthotropic materials[8] and for other kinds of transducers excitations.[9,10] Although these studies gave conclusive results, the recovery of the solutions in the time and space domains are complex since it requires evaluation of the inverse of the integral transforms.

Another approach was based on the normal-mode expansion method,[11] dealing with the problem of waveguide excitation. The excited acoustic field was expressed as the sum of the ''normal modes'' in the waveguide in this method. By using the orthogonality condition between the normal modes, the contribution of each normal mode to the excited acoustic field could then be determined. This method has already been successfully applied in various cases including harmonic excitations by solid[12] and liquid[13] wedge transducers, as well as Hertzian contact transducers[14] and Lamb waves excited by lasers.[15]

a)Electronic mail: sebastien.grondel@univ-valenciennes.fr

 0001-4966/2002/112(1)/84/7/$19.00

Numerical techniques can also be used to model Lamb wave excitation. Contrarily to the previous methods, these numerical techniques make it possible to analyze structures with complex geometry and to study in more detail the Lamb wave interaction with a defect. Thus, Guo *et al.*[16] used the finite element method (FEM) in order to characterize the interaction of Lamb waves with delaminations in composite plates. Similarly, a recent study performed by Lowe *et al.*[17] was devoted to the transmission of Lamb waves across adhesively bonded lap joints. Furthermore, some numerical codes allow direct use of the electric signal as initial condition. Moulin *et al.*[18] presented a study on piezoceramic transducers embedded in composite structures. A frequency was determined for which the number of Lamb modes could be reduced to one mode. The study included modeling of the transducer behavior.

Recently hybrid techniques have been developed using both analytical methods and numerical models to avoid the problem associated with the FE techniques, such as wave mode separation and computation time. Cho *et al.*,[19] for example, associated the boundary element method (BEM) with the normal-mode expansion method to study the edge reflection of Lamb waves. This hybrid method was also applied to characterize the Lamb wave generation with time-delay periodic linear array[20] and to estimate guided-wave mode conversion in structures with thickness variation.[21]

The objective was to develop a transducer which is adapted for health monitoring of orthotropic composite plates. The paper first states the problem, and further presents the theoretical study using a hybrid technique. An experimental investigation was further conducted to validate the theory as well as to evaluate the selective Lamb-mode technique (SLMT) for a selected mode in application to health monitoring of composites.

II. PROBLEM STATEMENT

A. Specimen description

The study focused on the excitation of Lamb waves by thin piezoceramic elements in composite plates made of carbon fiber reinforced epoxy prepreg. This material is referred as HTA/6376C. The laminate dimensions were $600\times400\times4.78\ \text{mm}^3$. The stacking sequence of the cross-ply laminates was $[0_4/90_4/0_4/90_4/0_2]_s$. The material data are given in the Appendix.

B. Choice of the excitation domain

The excitation frequency domain was defined as the frequency bandwidth centered at the excitation frequency. The selection of the excitation frequency domain followed three conditions. The first condition concerned the number of generated modes. The smallest number of Lamb modes should be generated to limit the complexity of the received signals. As a second condition the Lamb mode was chosen to be as nondispersive as possible to have the waveform not dependent on the propagation distance. The third condition was then to select a Lamb-mode wavelength that was lower than or equal to the size of the damage to be detected by the health monitoring system. This condition is necessary but not sufficient to insure the interaction with the damage.

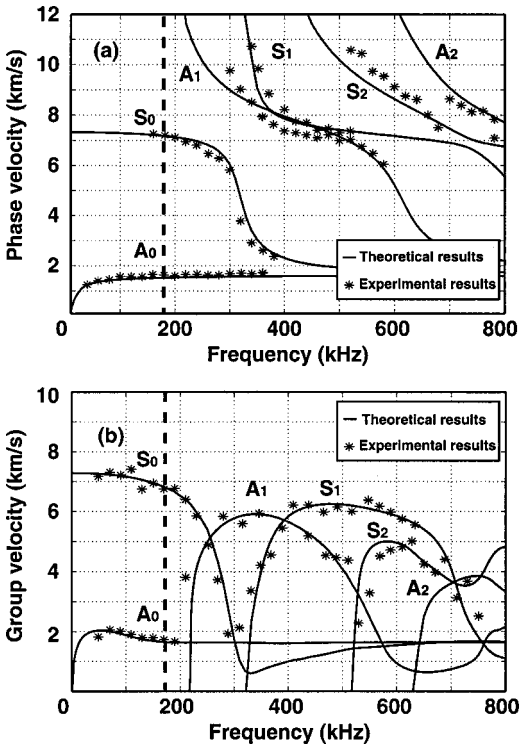


FIG. 1. Dispersion curves: theoretical (solid line) and experimental (dots); phase (a) and group (b) velocities.

Before applying those conditions to the transducer excitation signal, it is of primary importance to know the characteristics of the Lamb waves that can be generated in the plates. Figures 1(a) and (b) present the dispersion curves for the first six Lamb-modes propagating along the plate length.

The theoretical curves for the phase velocity [Fig. 1(a)] were determined using the formalism of Nayfeh *et al.*[22] The theoretical group velocity [Fig. 1(b)] was derived from the phase velocity.[6] Both group and phase velocities were further verified experimentally. The power spectral density[23] (PSD) and the cross-correlation method[24] were used to measure the phase and group velocities, respectively. The experimental and theoretical velocities were of good agreement, as shown in Figs. 1(a) and (b).

The dispersion curves of the group velocity show that at 180 kHz the modes S_0 and A_0 are close to be nondispersive, and that only those two modes can exist. Moreover, the group velocity of the S_0-mode is three times larger than that of the A_0-mode. Therefore, these two modes can be easily distinguished in the Lamb wave response. The wavelengths of the S_0- and A_0-modes at 180 kHz were 40 and 8.5 mm, respectively. The A_0-wavelength was about five times smaller than that of S_0. The A_0-mode was therefore expected to be more sensitive to damage detection. This initiated the interest of making the A_0-mode dominant.

C. Piezoceramic element description

A basic rule[1] in ultrasonic excitation is to excite a piezoceramic element at its natural resonance frequencies, in order to obtain the most efficient conversion of the electrical into

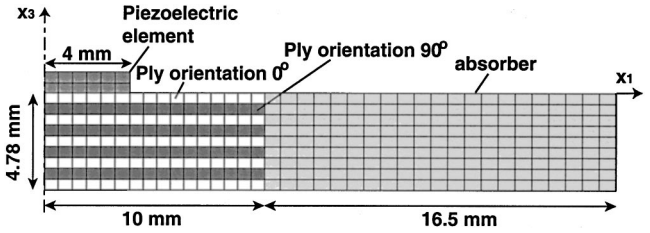


FIG. 2. Finite element model corresponding to the excitation of an 8-mm-element transducer.

mechanical energy. This leads to the choice of the piezoceramic element dimensions. The piezoceramic elements used in this investigation corresponded to the P1-60 type (quartz and silice) with a longitudinal dimension of 8 mm ($30\times8\times1\ \text{mm}^3$ in size) because they had a transverse vibration resonance at 180 kHz.

The theoretical study of the following section covered, first, the Lamb waves generated by such a transducer made of a single piezoceramic element at an excitation frequency of 180 kHz. Second, the transducer was extended to two piezoceramic elements: a dual-element transducer. The distance from the center of one piezoceramic element to the other was called the interelement distance. The investigation further focused on the selection of the smallest possible interelement distance between the two-piezoceramic elements in the dual-element transducer.

III. FEM AND NORMAL MODE ANALYSIS

The method described in this section is an extension of the hybrid modeling technique developed by Moulin *et al.*[25,26] This technique consisted of the normal-mode expansion method, which is suitable for waveguide excitation problems, and the finite element method, used to calculate the mechanical excitation applied by a piezoceramic transducer.

A. Finite element method (FEM)

A two-dimensional finite element model was performed using FE code ATILA,[27] as shown in Fig. 2. The model included a piezoceramic element as well as a part of the composite plate. This model included the composite by modelling the ply orientations in each group of 0°- and 90°-plies.

Due to symmetry conditions at the x_3 axis, one-half of the plate was modeled. The piezoceramic element was excited by a harmonic electric signal of angular frequency, ω, and of normalized amplitude in the model. Since the transverse dimension of the piezoceramic element was significantly larger than its longitudinal dimension and thickness, the FE model was reduced to two dimensions. The longitudinal dimension of the transducer was equal to $2L$, i.e., 8 mm. Furthermore, in the model, an ideal absorber was located at one end of the plate. It had similar mechanical properties as the plate, except the lossy part.[25] The loss coefficients of this lossy part were chosen as small as possible so that the standing wave due to the reflections from the composite/lossy-composite interface was negligible. They were, however, chosen large enough to sufficiently damp the modeled structure. The size[27] of each element in the finite element mesh was chosen to be at least smaller than $\lambda/4$, where λ is the smallest wavelength from all generated modes, to maintain a large accuracy of the FE results. The tangential and normal mechanical stresses (T^s_{13} and T^s_{33}, respectively) applied to the plate surface by the piezoceramic element were further calculated by FEM. Those stresses were subsequently substituted in the analytical model equation in order to determine the mode type and the amplitude of the generated Lamb waves.

B. Normal mode expansion method

Normal mode analysis suggests that any arbitrary displacement field $\mathbf{u}(x_1,x_3)$ and associated stress distribution $\mathbf{T}(x_1,x_3)$ can be expressed as a sum of the normal modes as[11]

$$\mathbf{u}(x_1,x_3)=\sum_m a_{\{m\}}(x_1)\mathbf{u}_{\{m\}}(x_3), \tag{1}$$

$$\mathbf{T}(x_1,x_3)=\sum_m a_{\{m\}}(x_1)\mathbf{T}_{\{m\}}(x_3), \tag{2}$$

where $a_{\{m\}}(x_1)$ are the normal mode amplitudes. $\mathbf{u}_{\{m\}}(x_3)$ and $\mathbf{T}_{\{m\}}(x_3)$ are the normal displacement and stress distributions, respectively.

For the model in Fig. 2, the following displacement and stress components were considered: u_1, u_3, T_{11}, T_{33}. Hence, the normal mode amplitude corresponding to the excitation force $f_{S\{m\}}(x_1)$ of a single-element transducer, applied in the interval $-L<x_1<L$, was given by

$$a_{\{m\}}(x_1)=\frac{e^{-j\xi_{\{m\}}x_1}}{4P_{\{mm\}}}\int_{-L}^{L} f_{S\{m\}}(\eta)e^{j\xi_{\{m\}}\eta}\,d\eta, \tag{3}$$

with

$$f_{S\{m\}}(x_1)=j\omega[u^*_{\{m\}1}(d)T^s_{13}(x_1,d)+u^*_{\{m\}3}(d)T^s_{33}(x_1,d)], \tag{4}$$

where $\xi_{\{m\}}$ is the wave number of the normal mode and $P_{\{mm\}}$ its corresponding average power flow per unit length along the x_1 direction. For a plate thickness of $2d$, the power flow is given by

$$P_{\{mm\}}=\frac{\omega}{2}\,\mathrm{Im}\left[\int_{-d}^{d}(u^*_{\{m\}1}T_{\{m\}11}+u^*_{\{m\}3}T_{\{m\}33})\,dx_3\right]. \tag{5}$$

For an excitation performed by N identical piezoceramic elements excited in phase, the solution was obtained by applying the superimposition principle. The stress state applied to the plate corresponded to the superimposition of all element stress of the multi-element transducer. The stresses in each element were assumed to be the same as those in the single-element transducer. The normal mode amplitude could then be rewritten as

$$a_{\{m\}}(x_1,p)=\left(\sum_{n=0}^{N-1} e^{j\xi_{\{m\}}np}\right)a_{\{m\}}(x_1), \tag{6}$$

where p corresponds to the interelement distance.

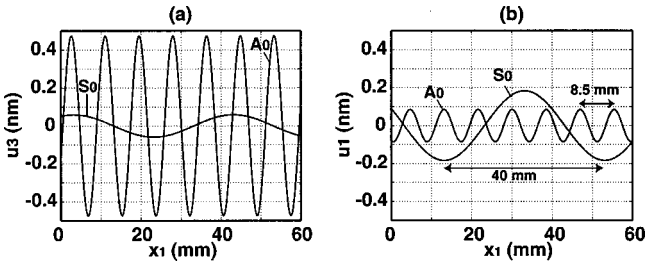


FIG. 3. Normal (a) and tangential (b) displacements at the plate surface of the modes S_0 and A_0 in terms of propagation distance, for an 8-mm-element transducer excited at 180 kHz.

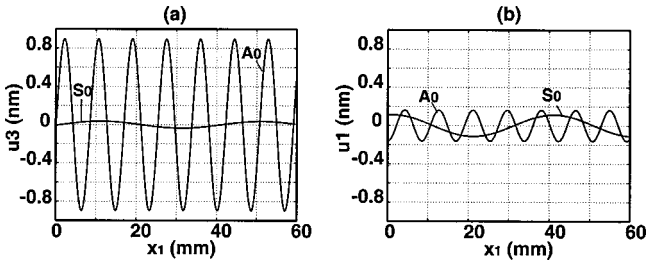


FIG. 5. Normal (a) and tangential (b) displacements at the plate surface of the modes S_0 and A_0 in terms of propagation distance, for a dual-element transducer excited at 180 kHz.

The influence of the interelement distance on the generated Lamb waves was determined by (6), and the results are discussed in the following section.

C. Theoretical results

The generation of Lamb waves was first investigated using a single piezoceramic element. The normal amplitude of each mode was determined by substituting the calculated surface stresses in (3). The displacement components u_3, u_1, at any location in the plate, were further obtained from (1) as shown in Figs. 3(a) and (b), respectively. The two modes A_0 and S_0, generated by a piezoceramic element excited at 180 kHz, were easily identified by their wavelengths. These figures illustrated the importance of the A_0-mode in the normal displacement of the plate and that of the S_0-mode in the tangential displacement. Neither of the two modes could be considered as dominant to each other since the difference in amplitude between the normal and tangential displacements was small.

To achieve the generation of the predominant A_0-mode, a double excitation by two piezoceramic elements was used. The size of each element was the same as those for a single-element transducer. Both elements of the dual-element transducer were excited in-phase at a frequency of 180 kHz. The theoretical displacement amplitude of the generated Lamb modes as a function of the interelement distance, *p*, was obtained from (6) as shown in Figs. 4(a) and (b). The displacement amplitude of a mode is maximal for a value of the interelement distance which is a multiple of the mode wavelength. For example, the amplitude of the A_0-mode is maximal for *p* equal to 17 mm, which is twice its corresponding wavelength. On other hand, the amplitude of the S_0-mode has a maximum at 40 mm.

Figures 5(a) and (b) illustrate the displacement of the Lamb wave modes at the plate surface for an interelement distance of 17 mm. The contribution of the S_0-mode to the normal displacement of the plate became, in that case, negligible to that of the A_0-mode. The tangential displacement amplitude of the A_0-mode has increased by a factor of 2 while that of S_0-mode has decreased also by a factor 2, to be level with the A_0-mode. The theoretical investigation demonstrated that the use of a dual-element transducer would make the A_0-mode significantly dominant.

IV. EXPERIMENTAL STUDY

A. Lamb-mode displacement measurement

In the test setup, all piezoceramic elements were surface-bonded with epoxy to the composite plate surface, as shown in Fig. 6. Those elements were excited by a toneburst of five-sinusoidal cycles of frequency 180 kHz and of amplitude ± 10 V. The toneburst was also windowed by a Hanning function. The Lamb wave receiver was a conventional surface mounted transducer (Panametrics A413-SB) coupled to the plate using coupling media. It was only used to measure the normal displacements. This conventional sensor operates within a large frequency bandwidth centered on 500 kHz.

The amplitudes of the generated Lamb-modes in the plate were determined using the two-dimensional Fourier transform method[28] for both single- and dual-element transducers. The dual-element transducer consisted of two elements (E1 and E2) that were excited in-phase with the optimized interelement distance of 17 mm to generate a dominant A_0-mode. The extrapolation of the harmonic excitation in the theoretical model to the toneburst for the experiments is only possible since the wave numbers of both the A_0- and S_0-modes were distinct from each other around the considered excitation frequency, and also because the frequency domain did not correspond to dispersive modes.

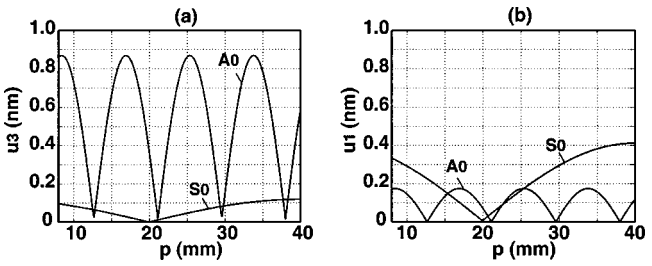


FIG. 4. Normal (a) and tangential (b) displacements at the plate surface of the S_0- and A_0-modes as a function of the interelement distance.

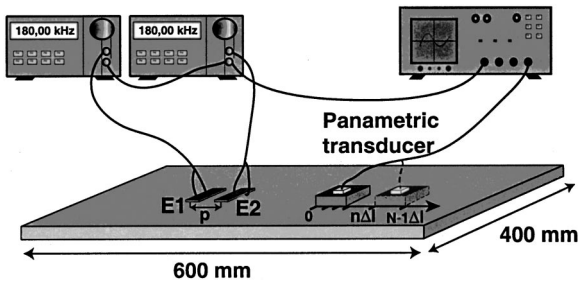


FIG. 6. Experimental setup for in-phase excitation of a dual-element transducer.

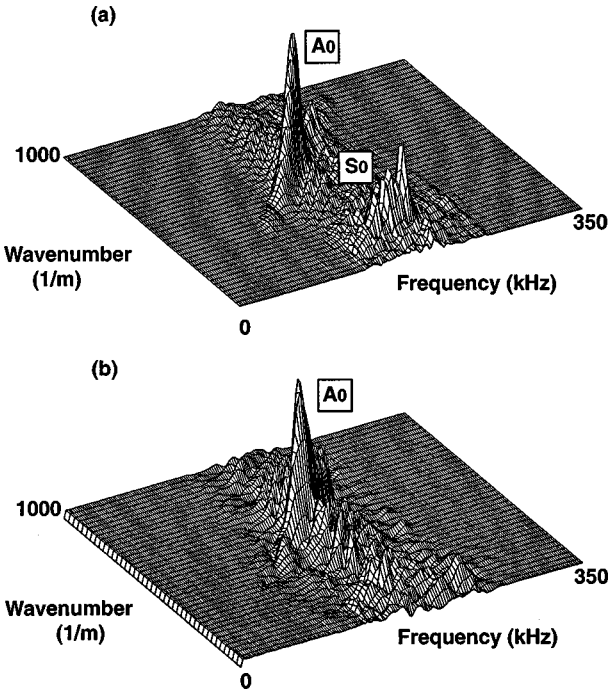


FIG. 7. 2D Fourier transform of generated Lamb waves by (a) single- and (b) dual-element transducers.

A series of 64 waveforms was recorded as the receiver was slid away from the generators with increments of 2 mm, along the longitudinal direction of the plate. Each Lamb wave response consisted of 1000 samples. The sampling serial of the experiments was 500 ns. Before measuring the 2D Fourier transform, zeros were padded to the end of the signal in both spatial and temporal domains with 64 and 24 zeros, respectively, in order to smooth the results. A 2D Fourier transform routine was processed to the data matrix resulting in a return matrix of spatial frequencies (wave numbers) versus temporal frequencies. This method enables the amplitude of different modes propagating at the same temporal frequencies to be determined.

Figures 7(a) and (b) give respectively the results of the 2D Fourier transform for both single- and dual-element transducers. In Fig. 7(a), two modes were clearly identified around 180 kHz, as the S_0- and A_0-modes. The amplitude of the A_0-mode was greater than that of the S_0-mode, which was in agreement with the results obtained from the theoretical model. The normal displacement amplitude of the A_0-mode was in fact twice larger than that of the S_0-mode.

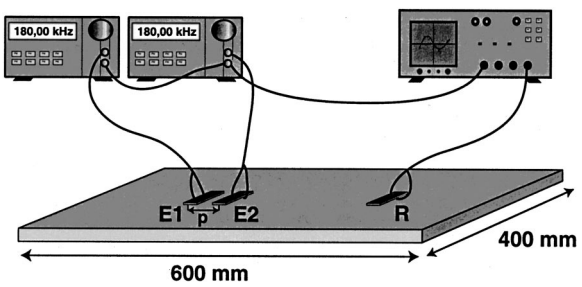


FIG. 8. Experimental setup for Lamb wave generation by a dual-element transducer [E1,E2] and reception by a single-element transducer R.

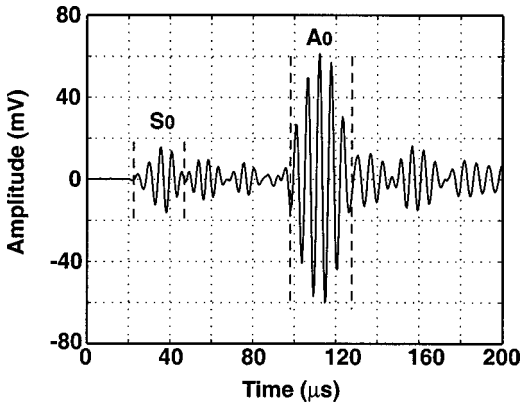


FIG. 9. Lamb wave response received by a 4-mm-element transducer.

This quantitatively agreed with the results from the theory, as indicated in Fig. 3(a).

The excitation with a dual-element transducer generated only the A_0-mode in agreement with the theory as shown in Fig. 7(b).

B. Lamb wave response filter

Previous works[6,29] demonstrated the amplitude of a given mode was maximal when the longitudinal dimension of the piezoceramic element generating the Lamb waves was equal to half of a mode wavelength. To the contrary, the amplitude of a mode was minimal for an element longitudinal dimension equal to the mode wavelength, called the integration phenomenon. Those considerations concerned the generation of Lamb waves, but this may also be applicable to Lamb wave reception.

A dual-element transducer constituted of two piezoceramic elements E1 and E2, as shown in Fig. 8, generated Lamb waves. The Lamb wave response was further obtained by the Lamb wave receiver, which was a surface bonded piezoceramic element, R. The longitudinal dimension of the piezoceramic element R was chosen to be 4 mm. This longitudinal dimension corresponds to the half wavelength of the A_0-mode, expecting the reception of the A_0-mode with the maximum amplitude. The R-element was about 150 mm in distance from the E2-element. The Lamb wave response contained the A_0-mode significantly dominating the S_0-mode in agreement with the theory, as shown in Fig. 9.

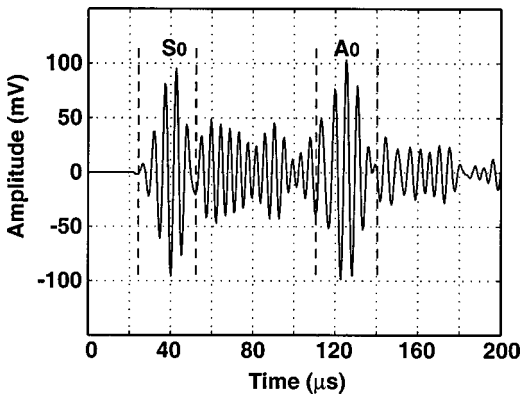


FIG. 10. Lamb wave response collected by an 8-mm-element transducer.

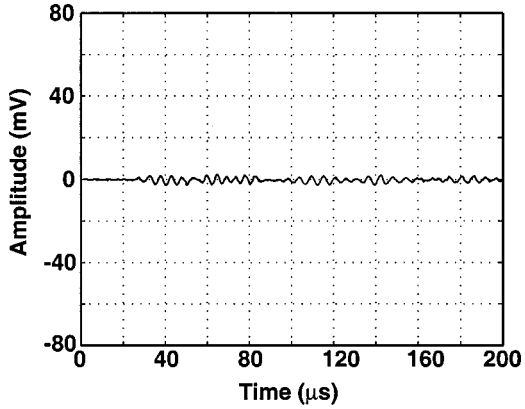


FIG. 11. Difference between the measured Lamb wave responses; before (a) and after (b) 1 J impact.

In another experiment, the Lamb wave receiver, R, was a 8-mm-element transducer, which was also surface bonded to the composite plate. The receiver longitudinal dimension was equal to the A_0-mode wavelength, under the integration phenomenon. The Lamb wave receiver was distant from the E2-element by 170 mm. As expected, the dominance of the A_0-mode was lost, as shown in Fig. 10.

The present results showed the influence of the transducer longitudinal dimension in the reception of selective Lamb-modes. On the other hand, such a Lamb wave receiver had a reduced sensitivity, with amplitude decrease in the Lamb wave response, since the Lamb wave receiver did not operate at its transverse resonance frequency which is 360 kHz for the 4-mm piezoceramic element.

C. An example of damage detection

The tests discussed in the previous section demonstrated the possibility of generating and receiving a given Lamb-mode, the A_0-mode, by the selection of appropriate transducers. The objective in this section was to use this improvement of the transducer in a damage monitoring system of composites. The composite plate was impacted by a low-velocity impactor. The impacts were performed at the center of the composite plate between the receiver and the generator. The impactor mass was 2.052 kg and the impactor head was of half-spherical shape 16 mm in diameter. The impact energy was obtained by changing the drop height.

As an initial test to demonstrate the robustness of the technique, an impact of 1 J energy was performed. At this low impact energy, no damage occurred as indicated from the C-scan examination. No significant changes were detected in the Lamb wave response as expected. The difference between the two Lamb wave responses, collected by the Lamb wave receiver before and after 1 J impact is presented in Fig. 11.

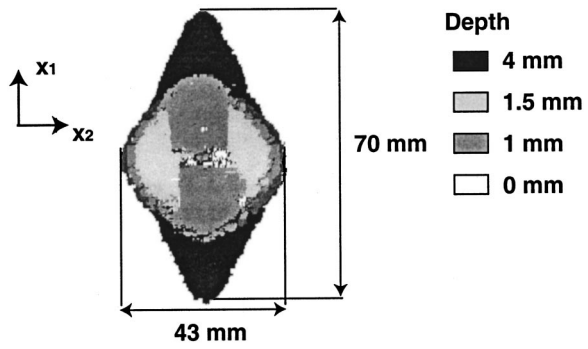


FIG. 12. C-scan of a 24 J impacted composite plate.

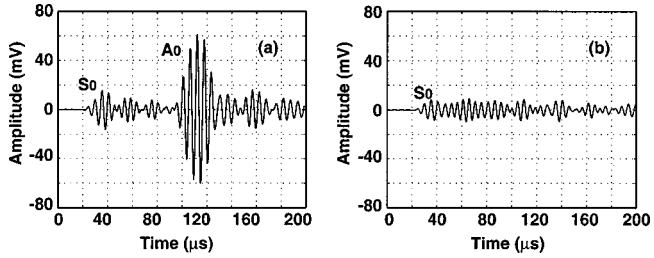


FIG. 13. Lamb wave responses collected from a 4-mm-element transducer; before (a) and after (b) 24 J impact.

The plate was impacted with an energy level of 24 J, expecting to generate an impact damage with delaminations through the entire thickness. An early study[30] showed that damage appeared at approximately 12 J. The C-scan image, given in Fig. 12, showed a damage size of approximately 3000 mm^2.

Figures 13(a) and (b) represent the Lamb wave responses collected by the 4-mm-element transducer, R, before and after impact, respectively. The amplitude of the A_0-mode decreased with the increasing impact energy, therefore revealing a large sensitivity of this mode towards a composite damage. This was expected since the A_0-mode has the smallest wavelength likely to interfere with the damage. This experiment demonstrated the advantage of the selective Lamb-mode technique (SLMT) for the damage monitoring of Lamb waves.

V. CONCLUSIONS

Theoretical and experimental investigations showed the possibility of designing transducers in such a way that the A_0-mode becomes dominant over the S_0-mode propagating in orthotropic composites. Great attention was paid to both Lamb wave generator and Lamb wave receiver, which have a direct effect on the Lamb wave response. This selective Lamb-mode technique (SLMT), based on the dominant A_0-mode, was then applied for damage detection in composites. The tests revealed a large sensitivity of this A_0-mode towards a composite damage, demonstrating the potential of the selective Lamb-mode technique for in-service structural health monitoring.

ACKNOWLEDGMENTS

The authors from the Swedish Defense Research Agency (FOI) would like to thank the Structures Department of FOI and the Swedish Defense Materiel Administration (FMV) for financial support. The other authors acknowledge the IEMN Laboratory, the CNRS, and the Valenciennes University for financial support.

APPENDIX: MATERIAL DATA OF A COMPOSITE PLY

The prepregs were manufactured by Hexcel (UK). The matrix was made of the epoxy referenced as Fibredux 6376C, designed for aerospace composites. The unidirectional fibers, along the x_1 axis, were the high tensile carbon fibres (HTA), produced by Enka. The x_3 axis was defined as the perpendicular axis to the ply-plane.

Typical ply for the HTA/6376:

Nominal ply thickness: 0.133 mm

Mass density: 1610 kg/m^3

Elastic tensor: $E_{11}=145$ GPa, $E_{22}=E_{33}=10.5$ GPa, $\nu_{12}=\nu_{13}=0.3$, $\nu_{23}=0.6$, $G_{12}=G_{13}=5.25$ GPa, $G_{23}=3.3$ GPa.

[1] S. Grondel, C. Delebarre, J. Assaad, J. P. Dupuis, and L. Reithler, "Fatigue crack monitoring of riveted aluminum strap joints by Lamb wave analysis and acoustic emission measurement techniques," NDT & E Int. **35**(3), 137–146 (2002).

[2] S. Grondel, "Contribution à l'optimisation du contrôle santé intégré par ondes de Lamb. Application à la surveillance de structures aéronautiques" ("Optimisation and development of a Health monitoring system based on Lamb waves. Application to the monitoring of aeronautic structures"), Ph.D. thesis, Report 0034, University of Valenciennes, France, Dec. 2000.

[3] Y. S. Roh and F.-K. Chang, "Effect of impact damage on Lamb wave propagation in laminated composites," Dynamic Response and Behavior of Composites ASME, AD-Vol. **46**, 127–138 (1995).

[4] N. Bourasseau, E. Moulin, C. Delebarre, and P. Bonniau, "Radome health monitoring with Lamb waves: experimental approach," NDT & E Int. **33**(6), 393–400 (2000).

[5] C. A. Paget and K. Levin, "Structural integrity of composites with embedded piezoelectric ceramic transducer," in *Smart Structures and Materials 1999: Smart Structures and Integrated Systems*, edited by M. Wereley, Proc. SPIE **3668**, 306–313 (1999).

[6] I. A. Viktorov, *Rayleigh and Lamb Waves* (Plenum, New York, 1967).

[7] J. Miklowitz, *The Theory of Elastic Waves and Waveguides* (North Holland, Amsterdam, 1978).

[8] S. P. Pelts and J. L. Rose, "Source influence parameters on elastic guided waves in an orthotropic plate," J. Acoust. Soc. Am. **99**, 2124–2129 (1996).

[9] J. L. Rose, S. P. Pelts, and M. J. Quarry, "A comb transducer model for guided wave NDE," Ultrasonics **36**, 163–169 (1998).

[10] M. Castaing and P. Cawley, "The generation, propagation, and detection of Lamb waves in plates using air-coupled ultrasonic transducers," J. Acoust. Soc. Am. **100**, 3070–3077 (1996).

[11] B. A. Auld, *Acoustic Fields and Waves in Solids*, 2nd ed. (Krieger, FL, 1990), Vol. 2.

[12] J. J. Ditri and J. L. Rose, "Excitation of guided waves in generally anisotropic layers using finite sources," Trans. ASME **61**, 330–338 (1994).

[13] X. Jia, "Modal analysis of Lamb wave generation in elastic plates by liquid wedge transducers," J. Acoust. Soc. Am. **101**, 834–842 (1997).

[14] F. L. Degertekin and B. T. Khury-Yakub, "Lamb wave excitation by Hertzian contacts with application in NDE," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 769–778 (1997).

[15] J. C. Cheng and S. Y. Zhang, "Quantitative theory for laser-generated Lamb waves in orthotropic thin plates," Appl. Phys. Lett. **74**(14), 2087–2089 (1999).

[16] N. Guo and P. Cawley, "The interaction of Lamb waves with delaminations in composite laminates," J. Acoust. Soc. Am. **94**, 2240–2246 (1993).

[17] M. J. S. Lowe, R. E. Challis, and C. W. Chan, "The transmission of Lamb waves across adhesively bonded lap joints," J. Acoust. Soc. Am. **107**, 1333–1345 (2000).

[18] E. Moulin, J. Assaad, and C. Delebarre, "Piezoelectric transducer embedded in a composite plate: Application to Lamb wave generation," J. Appl. Phys. **82**(5), 2049–2055 (1997).

[19] Y. Cho and J. L. Rose, "A boundary element solution for a mode conversion study on the edge reflection of Lamb waves," J. Acoust. Soc. Am. **99**(4), 2097–2109 (1996).

[20] W. Zhu and J. L. Rose, "Lamb wave generation and reception with time-delay periodic linear arrays: A BEM simulation and experimental study," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**(3), 654–664 (1999).

[21] Y. Cho, "Estimation of ultrasonic guided wave mode conversion in a plate with thickness variation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**(3), 591–603 (2000).

[22] A. H. Nayfeh and D. E. Chimenti, "The general problem of elastic waves propagation in multilayered anisotropic media," J. Acoust. Soc. Am. **89**, 1521–1531 (1991).

[23] S. Grondel, J. Assaad, C. Delebarre, P. Blanquet, and E. Moulin, "The propagation of Lamb waves in multilayered plates: Phase-velocity measurement," Meas. Sci. Technol. **10**, 348–353 (1999).

[24] S. M. Ziola and M. R. Gorman, "Source location in thin plates using cross-correlation," J. Acoust. Emiss. **9**(5), 2551–2556 (1991).

[25] E. Moulin, J. Assaad, C. Delebarre, and D. Osmont, "Modeling of Lamb waves generated by integrated transducers in composite plates using a coupled finite element-normal modes expansion method," J. Acoust. Soc. Am. **107**, 87–94 (2000).

[26] E. Moulin, J. Assaad, C. Delebarre, S. Grondel, and D. Balageas, "Modeling of integrated Lamb waves generation systems using a coupled element-normal modes expansion method," Ultrasonics **38**, 522–536 (2000).

[27] B. F. Hamonic, J. C. Debus, and J. N. Decarpigny, "The finite element code ATILA," Proceedings of the workshop, Toulon, France (1990).

[28] D. Alleyne and P. Cawley, "A two-dimensional Fourier transform method for the measurement of propagating multimode signals," J. Acoust. Soc. Am. **89**, 1159–1168 (1991).

[29] T. Demol, "Etude de transducteurs en barrette pour le contrôle santé de structures aéronautiques composites par ondes de Lamb. Application à la caractérisation de l'impact basse vitesse" ("Study and development of multi-element transducers adapted to the Health monitoring of aeronautic structures with Lamb waves. Application to impact characterisation"), Ph.D. thesis, Report 9802, University of Valenciennes, France, Jan. 1998.

[30] C. A. Paget, "Contribution au contrôle actif de santé de structures composites aéronautiques par transducteurs piézoélectriques insérés" ("Contribution to active health monitoring of aerospace composite structures by embedded piezoceramic transducers"), Ph.D. thesis, Report 0117, University of Valenciennes, France, June 2001.

Propagation of quasiplane nonlinear waves in tubes and the approximate solutions of the generalized Burgers equation

Michal Bednarik and Petr Konicek
Czech Technical University in Prague, FEE, Technicka 2, 166 27 Prague, Czech Republic

(Received 19 October 2001; revised 11 April 2002; accepted 4 May 2002)

This paper deals with using the generalized Burgers equation for description of nonlinear waves in circular ducts. Two new approximate solutions of the generalized Burgers equation (GBE) are presented. These solutions take into account the boundary layer effects. The first solution is valid for the preshock region and gives more precise results than the Fubini solution, whereas the second one is valid for the postshock (sawtooth) region and provides better results than the Fay solution. The approximate solutions are compared with numerical results of the GBE. Furthermore, the limits of validity of the used model equation are discussed with respect to boundary conditions and radius of a circular duct. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1488940]

PACS numbers: 43.25.Cb, 43.25.Jh [MFH]

I. INTRODUCTION

Propagation of nonlinear sound waves in waveguides represents a very interesting physical problem. In the case that nonlinear waves travel in a gas-filled waveguide, we can observe phenomena such as nonlinear distortion, nonlinear absorption, diffraction, lateral dispersion, boundary layer effects, etc. All these phenomena can be described by means of the complete system of the equations of the hydrodynamics, see, e.g., Refs. 1–4: the Navier–Stokes momentum equation, the continuity (mass conservation) equation, the heat transfer (entropy) equation and the state equations. Unfortunately, we have not known a general solution of this system of equations and numerical solutions bring many problems regarding stability of the solutions and their time consumption. Consequently, it is sensible to simplify the fundamental system of equations if we ignore some phenomena or consider some of them weak. This simplification leads to a derivation of model equations of nonlinear acoustics.

There are a number of papers which are devoted to various aspects concerning propagation of nonlinear waves in waveguides. The viscous and thermal dissipative effects on the nonlinear propagation of plane waves in hard-walled ducts are treated for instance in the papers.[5–7] The authors deal with the dependence of the frequency on the dissipative and dispersive effects induced by the acoustic boundary layer. Experimental results focused on propagation of finite-amplitude plane waves in circular ducts are presented in Refs. 8 and 9. Here is demonstrated a very good agreement between experimental data and results obtained by means of the Rudnick decay model for the fundamental harmonic. Burns[10] obtained a fourth-order perturbation solution for finite-amplitude waves. However, his expansion breaks down for large times because it contains secular terms. He took into account dissipation but he neglected the mainstream dissipation with respect to the boundary dissipation. Keller and Millmann[11] found the solution of the model equation for inviscid isentropic fluids where they used a perturbation expansion adapted to eliminate secular terms and determined the nonlinear wave number shift for dispersive modes. Keller[12] utilized the results from Ref. 11. He rewrote the results in a form that is useful near the cutoff frequency, in order to show that the cutoff frequencies and resonant frequencies of modes in acoustic waveguides of finite length depend upon the mode amplitude. Nayfeh along with Tsai[13] presented the nonlinear effects of the gas motion as well as the lining nonlinear acoustic material properties on the wave propagation and attenuation in circular ducts. They obtained a second-order uniformly valid expansion by using the method of multiple scales. These authors presented work,[14] where they investigated nonlinear propagation in a rectangular duct whose side walls were acoustically treated by means of the method of multiple scales as well. Also Ginsberg[15] dealt with the nonlinear propagation in the rectangular ducts. He determined by an asymptotic method the nonlinear two-dimensional acoustic waves that occur within a rectangular duct of semi-infinite length as the result of periodic excitation. In the paper[16] he utilized the perturbation method of renormalization to study the effect nonlinearity on a hard-walled rectangular waveguide. Nonlinear wave interaction in a rectangular duct was investigated by Hamilton and TenCate in Refs. 17 and 18 as well. Multiharmonic excitation of a hard-walled circular duct was treated by Nayfeh in the work.[19] He used the method of multiple scales to derive a nonlinear Schrödinger equation for the temporal and spatial modulation of the amplitudes and the phases of waves propagating in a hard-walled duct. Foda presented his work in Ref. 20 which is concerned with the nonlinear interactions and propagation of two primary waves in higher order modes of a circular duct each at an arbitrary different frequency and finite amplitude. He used the renormalization method to annihilate secular terms in the obtained expression. If we take into account no diffraction effects, we can use for description of nonlinear plane waves in circular ducts the generalized Burgers equation. It is the Burgers equation which is supplemented by the term which represents boundary layer effects, see Refs. 21–25. The generalized Burgers equation enables to describe dissipative and dispersive effects that are caused by the boundary layer. Asymptotical and numerical solutions of this equation were presented by Sugimoto.[26] The approximate solution of the Burgers equation in the preshock region

without taking into account the boundary layer effects was presented, e.g., by Blackstock.[27] In the case that only weak diffraction effects are considered, it is possible to use the KZK equation, see Refs. 22, 25, and 28. However, the boundary layer is incorporated into the boundary condition in Ref. 22 or 25.

This paper deals with possibilities of using the generalized Burgers equation for description of nonlinear waves in circular ducts. There are presented new approximate solutions of the generalized Burgers equation for a harmonic excitation in both preshock and postshock region. Further, results of numerical solutions of the model equation and the analytic approximation solutions are compared. Finally, the limits of validity of the used model equation are discussed with respect to boundary conditions and radius of a circular duct.

II. GENERALIZED BURGERS EQUATION

If we assume the wall friction in the case that nonlinear waves propagate in a rigid tube, then a thin boundary layer appears near the tube wall. Within the boundary layer, the velocity component in the direction of the tube axis decreases from a mainstream value to zero at the tube wall. The boundary layer affects the acoustic waves not only near the walls but even in the entire volume. The boundary layer causes both energy dissipation and wave dispersion.

According to the second-order nonlinear theory we can obtain the generalized Burgers equation (GBE) which enables to describe weakly nonlinear waves in thermoviscous fluids when the boundary layer is assumed to have a small displacement effect on the mainstream[4]

$$\frac{\partial v}{\partial z}-\frac{\beta}{c_0^2}v\frac{\partial v}{\partial \tau}+\sqrt{2}B\frac{\partial^{1/2}v}{\partial \tau^{1/2}}=\frac{b}{2\rho_0 c_0^3}\frac{\partial^2 v}{\partial \tau^2}, \tag{1}$$

where v is the particle velocity, z is the coordinate along the axis of the tube, $\tau=t-z/c_0$ is retarded time, t is time, c_0 is the small-signal sound speed, b is the dissipation coefficient of the medium, ρ_0 is the ambient density, γ is the adiabatic index, equal to the ratio of the specific heat at constant pressure c_p to that at constant volume c_V, $\beta=(\gamma+1)/2$ is the coefficient of nonlinearity. The coefficient B corresponds with the boundary layer and is given as

$$B=\sqrt{\frac{\nu}{2c_0^2R_0^2}}\left(1+\frac{\beta_T c_0^2}{c_p\sqrt{\mathrm{Pr}}}\right), \tag{2}$$

where R_0 is the tube radius, ν is the kinematic viscosity, Pr is the Prandtl number, and β_T is the thermal expansion coefficient of fluid

$$\beta_T=-\frac{1}{\rho_0}\left(\frac{\partial \rho}{\partial T}\right)_{p,T=T_0}, \tag{3}$$

where T is temperature. If we suppose a perfect gas then

$$\frac{\beta_T c_0^2}{c_p}=\gamma-1. \tag{4}$$

The fractional derivative of the order 1/2 in Eq. (1) represents the following integrodifferential operator:

$$\frac{\partial^{1/2}v(z,\tau)}{\partial \tau^{1/2}}=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\tau}\frac{\partial v(z,\tau')}{\partial \tau'}\frac{d\tau'}{\sqrt{\tau-\tau'}}. \tag{5}$$

The third term in the left part of Eq. (1) represents the boundary layer effects and is valid on the condition that

$$\delta\ll R_0, \quad \delta\ll\lambda, \tag{6}$$

where δ is the boundary layer thickness

$$\delta\simeq\sqrt{\frac{\nu}{\omega}}, \tag{7}$$

where ω is an angular frequency.

The solution of the linearized GBE for the boundary condition $v(0,\tau)=v_m\sin(\omega\tau)$ can be expressed as

$$v=v_m\exp[-(\alpha+\alpha_b)z]\sin(\omega\tau-\alpha_b z), \tag{8}$$

where α is the attenuation coefficient for the classical thermoviscous loss mechanism

$$\alpha=\frac{b\omega^2}{2\rho_0 c_0^3} \tag{9}$$

and the attenuation coefficient α_b represents the losses due to the wall friction

$$\alpha_b=B\sqrt{\omega}. \tag{10}$$

The ration of both attenuation coefficients can be written as

$$\frac{\alpha}{\alpha_b}\simeq\frac{\delta}{\lambda}\frac{R_0}{\lambda}. \tag{11}$$

It is obvious that $\delta/\lambda>1$ for high frequencies but $R_0/\lambda\gg1$ because the first of the conditions (6) has to be satisfied. It means that the classical thermoviscous losses dominates for high frequencies in comparison to the boundary layer losses. Consequently, the condition $\delta/\lambda\ll1$ cannot be satisfied for higher wave form harmonics.

III. APPROXIMATION SOLUTIONS OF THE GBE

We can write Eq. (1) in the following nondimensional form:

$$\frac{\partial V}{\partial \sigma}-\frac{1}{2}\frac{\partial V^2}{\partial \theta}+\sqrt{2}D_0\frac{\partial^{1/2}V}{\partial \theta^{1/2}}=\frac{1}{G_0}\frac{\partial^2 V}{\partial \theta^2}, \tag{12}$$

where

$$\theta=\omega\tau, \quad \sigma=\frac{\beta v_m\omega}{c_0^2}z, \quad V=\frac{v}{v_m},$$
$$D_0=\frac{Bc_0^2\sqrt{\omega}}{\omega v_m\beta}, \quad G_0=\frac{2\beta v_m\rho_0 c_0}{b\omega}. \tag{13}$$

A. Preshock region

If we suppose the boundary condition

$$V(0,\theta)=\sin(\theta), \tag{14}$$

then we can seek the solution of Eq. (12) in the preshock region $0<\sigma<1$ in the following form, see Ref. 29:

$$V(\sigma,\theta)=A(\sigma)R(\sigma,\theta)=A(\sigma)\sin[\theta+\sigma A(\sigma)R(\sigma,\theta)]. \tag{15}$$

Using Eq. (15) we obtain

$$\frac{\partial V}{\partial \sigma}=\frac{\partial (AR)}{\partial \sigma}$$

$$=R\frac{\partial A}{\partial \sigma}+\frac{\cos(\theta+\sigma AR)\left(A+\sigma\dfrac{\partial A}{\partial \sigma}\right)AR}{1-\sigma A\cos(\theta+\sigma AR)}, \tag{16}$$

$$\frac{\partial V}{\partial \theta}=A\cos(\theta+\sigma AR)\left(1+\sigma A\frac{\partial R}{\partial \theta}\right)$$

$$\rightarrow \frac{\partial V}{\partial \theta}=\frac{A\cos(\theta+\sigma AR)}{1-\sigma A\cos(\theta+\sigma AR)}, \tag{17}$$

$$\frac{1}{G_0}\frac{\partial^2 V}{\partial \theta^2}=-\frac{A}{G_0}\sin(\theta+\sigma AR)\left(1+\sigma AR\frac{\partial R}{\partial \theta}\right)^2$$

$$+\frac{\sigma A^2}{G_0}\frac{\partial^2 R}{\partial \theta^2}\cos(\theta+\sigma AR). \tag{18}$$

Provided that $1/G_0\sim D_0\sim \mu$ we can suppose $\sigma A\ll 1$. If we ignore terms which are higher than the second-order terms, then we can simplify Eqs. (16), (17), and (18)

$$\frac{\partial V}{\partial \sigma}\simeq R\frac{\partial A}{\partial \sigma}+\frac{A^2R\cos(\theta+\sigma AR)}{1-\sigma A\cos(\theta+\sigma AR)}, \tag{19}$$

$$\frac{1}{G_0}\frac{\partial^2 V}{\partial \theta^2}\simeq -\frac{1}{G_0}V \tag{20}$$

or

$$\frac{\partial V}{\partial \theta}\simeq -\int V\,d\theta. \tag{21}$$

If we integrate Eqs. (20) and (21), we suppose that the integration constants are equal to zero with respect to the boundary condition (14). With the help of the expressions (19), (20), and (21), we can modify Eq. (12),

$$\frac{V}{A}\frac{dA}{d\sigma}+\frac{1}{G_0}V=-\sqrt{2}D_0\frac{\partial^{1/2}V}{\partial \theta^{1/2}}. \tag{22}$$

We can seek the approximate solution of Eq. (12) in the form of the modified Fubini solution, valid only in the preshock region $0<\sigma\ll 1$,

$$V=\mathrm{Im}\left\{\sum_{n=1}^{\infty}\frac{2J_n[n\sigma A_n(\sigma)]}{n\sigma}\exp[j(n\theta+\Phi_n(\sigma))]\right\}, \tag{23}$$

where $j=\sqrt{-1}$ and $A_n(\sigma)$ is the function $A(\sigma)$ for the given harmonic because Eq. (22) is a linear equation, thus it is valid for each harmonic separately. We can assume that dispersion effects are very small in this region, therefore we can describe them by means of a function $\Phi_n(\sigma)$. This function can be expressed on the basis of the solution for the linearized GBE (8),

$$\Phi_n(\sigma)=-D_0\sqrt{n}\sigma. \tag{24}$$

With the help of Eq. (21), we can express Eq. (5) as

$$\frac{\partial^{1/2}V}{\partial \theta^{1/2}}=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\theta}\frac{\partial V(\sigma,\theta')}{\partial \theta'}\frac{d\theta'}{\sqrt{\theta-\theta'}}$$

$$\simeq\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\theta}\frac{-\int V(\sigma,\theta')d\theta'}{\sqrt{\theta-\theta'}}d\theta'. \tag{25}$$

Substituting (23) into Eq. (25) we get

$$\frac{\partial^{1/2}V}{\partial \theta^{1/2}}=-\mathrm{Im}\left\{\sum_{n=1}^{\infty}\left[\frac{2J_n[n\sigma A_n(\sigma)]}{n\sigma}\frac{\exp[j\Phi(\sigma)]}{\sqrt{\pi}}\right.\right.$$

$$\left.\left.\times\int_{-\infty}^{\theta}\frac{\exp(jn\theta')}{jn\sqrt{\theta-\theta'}}d\theta'\right]\right\}. \tag{26}$$

We can solve the integral in Eq. (26),

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\theta}\frac{\exp(jn\theta')}{jn\sqrt{\theta-\theta'}}d\theta'=-\frac{\sqrt{jn}}{n^2}\exp(jn\theta)$$

$$=-\frac{1+j}{\sqrt{2n^3}}\exp(jn\theta). \tag{27}$$

Substituting Eq. (23) into Eq. (22) we obtain, with the help of expression (27), this equation for the function $A_n(\sigma)$

$$\frac{dA_n}{d\sigma}+A_n\left(\frac{1}{G_0}+\frac{1}{\sqrt{n^3}}D_0\right)=0. \tag{28}$$

Here dispersion effects are not taken into account because these effects are incorporated in Eq. (23), see Ref. 29. Solving Eq. (28) we get

$$A_n(\sigma)=C\exp\left[-\sigma\left(\frac{1}{G_0}+n^{-3/2}D_0\right)\right], \tag{29}$$

where C is an integration constant. For the case that $G_0\rightarrow\infty$ and $D_0\rightarrow 0$ the solution (23) has to represent the Fubini solution therefore the integration constant $C=1$.

Using (23), (24), and (29) we can write the approximate solution of Eq. (12) in the region $0<\sigma<1$ in this form

$$V(\sigma,\theta)=\mathrm{Im}\left[\sum_{n=1}^{\infty}a_n\exp(jn\theta-D_0\sqrt{n}\sigma)\right], \tag{30}$$

where

$$a_n=\frac{2J_n\left\{n\sigma\exp\left[-\sigma\left(\dfrac{1}{G_0}+n^{-3/2}D_0\right)\right]\right\}}{n\sigma}. \tag{31}$$

B. Postshock region

If we take into account only the boundary layer effects then Eq. (12) can be written as

$$\frac{\partial V}{\partial \sigma}-\frac{1}{2}\frac{\partial V^2}{\partial \theta}+\sqrt{2}D_0\frac{\partial^{1/2}V}{\partial \theta^{1/2}}=0. \tag{32}$$

We can seek the solution of Eq. (32) in the postshock region $\sigma\gg 3$ (after formation of a shock wave profile) in the following form, see Ref. 29:

$$V(\sigma,\theta)=\frac{A(\sigma)(\pi-\theta)}{1+\sigma A(\sigma)}=A(\sigma)R(\sigma,\theta)$$

$$\text{for } \theta\in\langle 0;2\pi\rangle. \tag{33}$$

The solution (33) applied to Eq. (32) yields

$$\frac{\partial}{\partial\sigma}(AR)-A^2R\frac{\partial R}{\partial\theta}=-\sqrt{2}D_0\frac{\partial^{1/2}V}{\partial\theta^{1/2}}. \tag{34}$$

After differentiating in Eq. (34) we obtain

$$\frac{V}{A}\frac{dA}{d\sigma}-\frac{\sigma(dA/d\sigma)}{1+\sigma A}V=-\sqrt{2}D_0\frac{\partial^{1/2}V}{\partial\theta^{1/2}}. \tag{35}$$

Fourier analysis of the solution (33) allows us to follow the behavior of the individual harmonics

$$V(\sigma,\theta)=\mathrm{Im}\left(\sum_{n=1}^{\infty}\frac{2}{n}\frac{A_n(\sigma)}{1+\sigma A_n(\sigma)}\exp(jn\theta)\right), \tag{36}$$

where $A_n(\sigma)$ is the function $A(\sigma)$ for the given harmonic because Eq. (35) is a linear equation with respect to the function V, thus it is valid for each harmonic separately.

Substituting the spectral solution (36) in Eq. (35), we have

$$\sum_{n=1}^{\infty}\left[\frac{2}{n}p_n\frac{1}{A_n}\frac{dA_n}{d\sigma}-\frac{2}{n}p_n\frac{\sigma\dfrac{dA_n}{d\sigma}}{1+\sigma A_n}+\sqrt{2}D_0\frac{\partial^{1/2}}{\partial\theta^{1/2}}\left(\frac{2}{n}p_n\right)\right]=0, \tag{37}$$

where

$$p_n=\frac{A_n\exp(jn\theta)}{1+\sigma A_n}. \tag{38}$$

From Eq. (37) we can get the equations that enable us to evaluate the complex Fourier coefficients

$$\frac{1}{A_n}\frac{dA_n}{d\sigma}-\frac{\sigma(dA_n/d\sigma)}{1+\sigma A_n}+\sqrt{j2n}D_0=0. \tag{39}$$

After rewriting Eq. (39), we have

$$\frac{dA_n}{d\sigma}+\sqrt{j2n}D_0\sigma A_n^2+\sqrt{j2n}D_0A_n=0, \tag{40}$$

where Eq. (40) represents the Bernoulli differential equation. If we want to corporate dispersion effects, it is necessary to regard a function A_n as a complex function.

The general solution of Eq. (40) can be expressed in the following form:

$$A_n=\frac{\exp(-\sigma\sqrt{j2n}D_0)}{C-\sigma\exp(-\sigma\sqrt{j2n}D_0)\left(1+\dfrac{1}{\sigma\sqrt{j2n}D_0}\right)}, \tag{41}$$

where C is an integration constant. It is necessary to choose the constant C in order to the coefficient $A_n=1$ for the case when $D_0\to 0$. That is,

$$C=1+\frac{1}{\sqrt{j2n}D_0}. \tag{42}$$

The constant C chosen by this way enables to reduce the solution (36) to the well-known solution for ideal fluids.

Substitution of the constant (42) into the solution (41) yields

$$A_n=\frac{\exp(-\sigma\sqrt{j2n}D_0)}{1-\sigma\left[\exp(-\sigma\sqrt{j2n}D_0)+\dfrac{\exp(-\sigma\sqrt{j2n}D_0)-1}{\sigma\sqrt{j2n}D_0}\right]}. \tag{43}$$

Let us consider the Burgers equation

$$\frac{\partial V}{\partial\sigma}-\frac{1}{2}\frac{\partial V^2}{\partial\theta}=\frac{1}{G_0}\frac{\partial^2V}{\partial\theta^2}. \tag{44}$$

If $G_0\gg 1$ and $\sigma\geqslant 3$ we can use the approximate solution of Eq. (44) which is known as the Fay's solution

$$V(\sigma,\theta)=\sum_{n=1}^{\infty}\frac{2}{G_0}\frac{\sin(n\theta)}{\sinh\left[\dfrac{n(1+\sigma)}{G_0}\right]}. \tag{45}$$

If $n(1+\sigma)/G_0<1$ is small it is possible to use the following substitution in the Fay's solution (45):

$$\sinh\left[\frac{n(1+\sigma)}{G_0}\right]\approx\frac{n(1+\sigma)}{G_0}. \tag{46}$$

Substitution of the approximation (46) in the Fay's solution yields

$$V(\sigma,\theta)=\sum_{n=1}^{\infty}\frac{2}{n}\frac{\sin(n\theta)}{1+\sigma}, \tag{47}$$

which is the approximate solution of the Burgers equation for ideal fluids.

Now, let us suppose the following approximate solution of the generalized Burgers equation Eq. (12):

$$V(\sigma,\theta)=\mathrm{Im}\left(\sum_{n=1}^{\infty}\frac{2}{G_0}\frac{A_n\exp(jn\theta)}{\sinh\left[\dfrac{n(1+\sigma A_n)}{G_0}\right]}\right). \tag{48}$$

Provided that $G_0\gg 1$ and $D_0\ll 1$ then it is possible to use this simplification

$$\frac{A_n}{\sinh\left[\dfrac{n(1+A_n\sigma)}{G_0}\right]}\approx\frac{A_n}{\dfrac{n(1+\sigma A_n)}{G_0}}. \tag{49}$$

If we use the approximation (49) in the solution (48), we obtain the solution (36).

At high frequencies the classical thermoviscous loss mechanism is dominant with respect to the wall losses (the classical attenuation $\sim\omega^2$, whereas the boundary layer attenuation $\sim\sqrt{\omega}$). In addition, at high frequency boundary layer dispersive effects are small enough and allow to approximate the phase speed c_{ph} by c_0. This follows from the relation

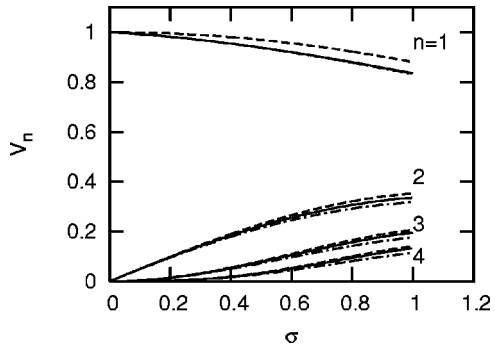


FIG. 1. Comparison of the preshock region approximate solution (solid lines), the Fubini solution (dashed lines), and the numerical solution (dashed–dotted lines) for the different harmonics, $G_0=50$, $D_0=0.05$.

$$c_{\mathrm{ph}}=\frac{c_0}{1+\dfrac{c_0 B'}{\sqrt{\omega}}}. \tag{50}$$

If we take into account the results mentioned above, we can assume that the solution (48) is the approximate solution of the generalized Burgers equation (12) in the region $\sigma \geqslant 3$ when $D_0 \ll 1$.

IV. COMPARISON OF APPROXIMATE SOLUTIONS

In this section we deal with comparison between the approximate (analytic) and numerical solutions of the generalized Burgers equation. The accuracy of the analytic solutions are investigated below for both the preshock and postshock region.

The Burgers equation (12) was solved by means of the standard Runge–Kutta method of the fourth order in the frequency domain (the first 100 harmonics were used). The numerical oscillations were damped by the method described by Fenlon[30] in the postshock region. Each harmonic was multiplied by the coefficient Ψ_n given by

$$\Psi_n=\frac{\sin(nH)}{nH}, \tag{51}$$

where H is the frequency damping coefficient. 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 in the preshock region.

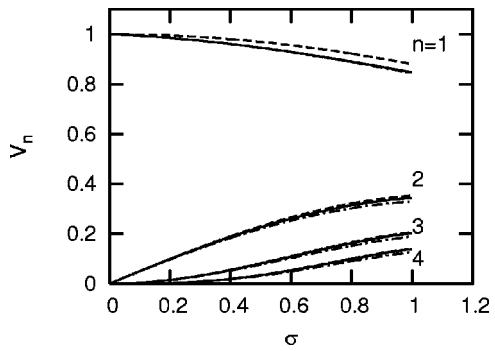


FIG. 2. Comparison of the preshock region approximate solution (solid lines), the Fubini solution (dashed lines), and the numerical solution (dashed–dotted lines) for the different harmonics, $G_0=500$, $D_0=0.05$.

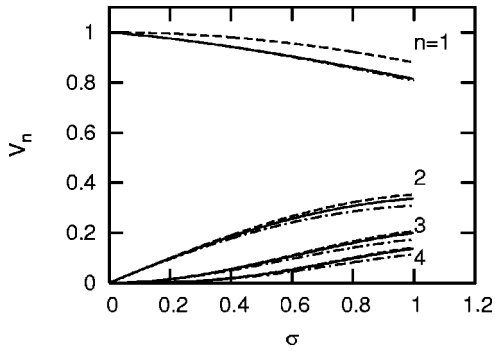


FIG. 3. Comparison of the preshock region approximate solution (solid lines), the Fubini solution (dashed lines), and the numerical solution (dashed–dotted lines) for the different harmonics, $G_0=500$, $D_0=0.1$.

A. Preshock region

Spatial evolution curves for the first four harmonic amplitudes are shown in Figs. 1–3. To illustrate the accuracy of the analytic solution (30) we choose the Goldberg number G_0 that equals 50 and 500 and the value of parameter D_0, which represents the boundary layer effects, equal to 0.05 and 0.1. In order to demonstrate the advantage of the analytic solution we also depicted the Fubini solution in the figures. As shown in Figs. 1 and 2, the analytic predictions of the fundamental and second harmonic amplitudes are in good agreement with the numerical results. It is obvious that the accuracy of the Fubini solution is worse than in the case of the analytic one. As it is seen in these figures the accuracy of the analytic solution becomes worse for lower values of G_0 but it is better than the accuracy of the Fubini solution. In Fig. 3 spatial evolution curves for the harmonics are depicted

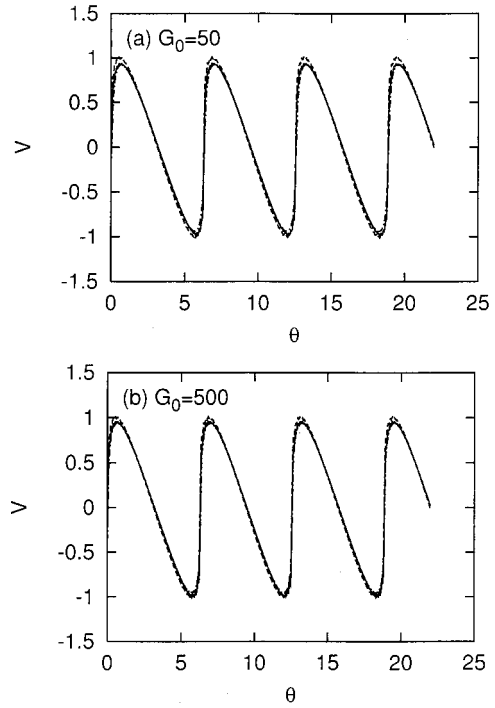


FIG. 4. Comparison of the preshock region approximate solution (solid lines), the Fubini solution (dashed lines), and the numerical solution (dashed–dotted lines) in the time domain, $D_0=0.05$, $\sigma=0.99$, (a) $G_0=50$, (b) $G_0=500$.

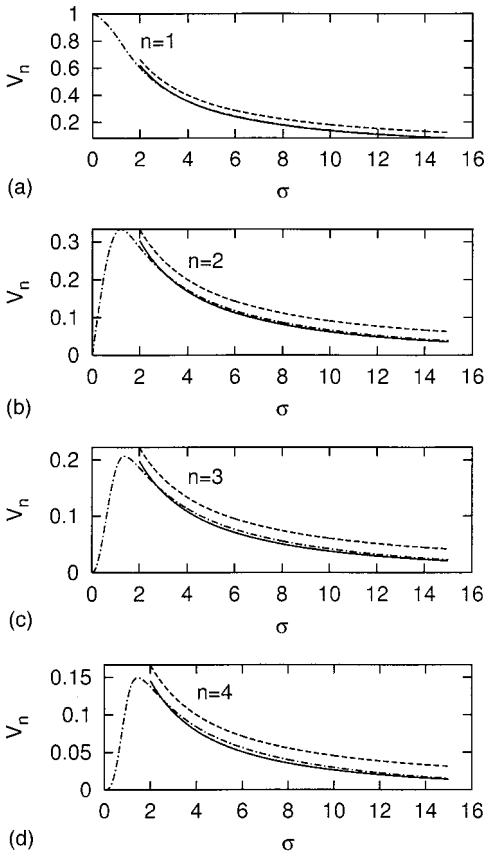


FIG. 5. Comparison of the postshock region approximate solution (solid lines), the Fay solution (dashed lines), and the numerical solution (dashed–dotted lines) for the different harmonics, $G_0=500$, $D_0=0.05$, (a) $n=1$, (b) $n=2$, (c) $n=3$, (d) $n=4$.

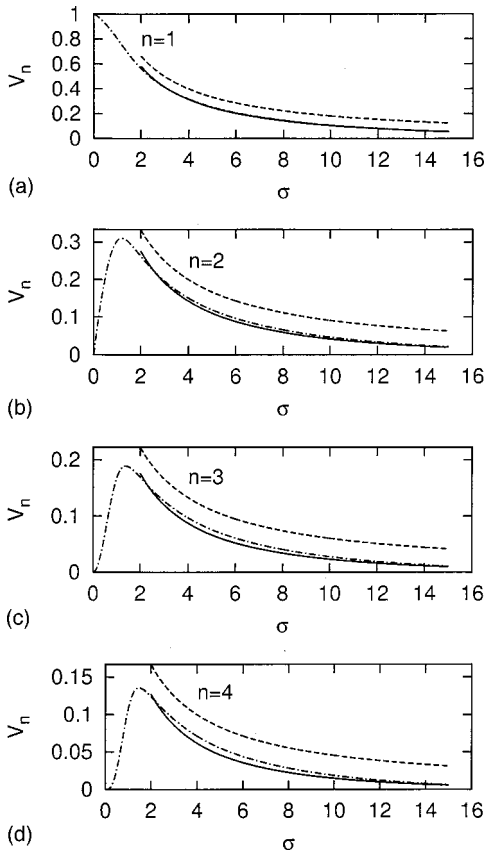


FIG. 6. Comparison of the postshock region approximate solution (solid lines), the Fay solution (dashed lines), and the numerical solution (dashed–dotted lines) for the different harmonics, $G_0=500$, $D_0=0.1$, (a) $n=1$, (b) $n=2$, (c) $n=3$, (d) $n=4$.

when $D_0=0.1$. From comparison of the three solutions follows that the analytic prediction of the fundamental harmonic is in good agreement with the numerical solution but the Fubini solution failed.

The wave forms are shown in Figs. 4(a) and (b) for the different Goldberg number ($G_0=50$ and 500) and $D_0=0.05$ at the distance $\sigma=0.99$. We can say on the basis of shown figures that the differences between the presented wave forms are very small, consequently we can consider sufficiently accurate both the analytic and Fubini solutions for the parameters G_0 and D_0 in the range which guarantees the validity of the analytic solution. We can see in Figs. 4(a) and (b) that the wave form asymmetry is not almost observable for small values of D_0.

B. Postshock region

Both spatial evolution curves for the first four harmonic amplitudes and wave forms are depicted in the postshock region (behind the shock formation distance) in Figs. 5 and 6 and Figs. 7 and 8. The comparisons are made between the numerical solution, the Fay's solution and the analytic one (48). As shown in the figures, the Fay's solution disagrees with the numerical one when the boundary layer effects play a dominant role, that is in the case that the parameter D_0 is large relatively. In contrast with the Fay's solution, the analytic solution is similar to the numerical one as we can see from the figures. In addition, the wave forms obtained from the analytic solution, which takes also into account the boundary layer effects, are asymmetric. The wave form asymmetry is caused by a boundary layer dispersion and occurs behind the shock formations because a phase shift between harmonics depends on a distance from the source. For the relatively large values of D_0 we can observe that there is a slight phase shift between the analytic and numerical solution, see Figs. 7 and 8. However, if values of D_0 are small then the agreement is excellent between both the solutions. As it is seen in the presented figures, the postshock analytic predictions are in good agreement with the numerical results even for higher values of D_0 contrary to the preshock analytic solution. Therefore the accuracy of the postshock analytic solution is better than the preshock analytic one.

There is a relatively good agreement between the Fubini solution and numerical one for small values of D_0, however the Fay's solution apparently disagrees with the numerical one. Thus the advantage of the postshock analytic solution is evident in contrast to the Fay's solution.

V. CONCLUSION

This paper presents the derivation of the new approximate solution of the generalized Burgers equation in the preshock region $0<\sigma<1$ for $1/G_0 \sim D_0 \sim \mu$. This approximate solution gives more precise results than the Fubini solution in the limits of its validity. The approximate solution converges to the Fubini solution when $G_0 \to \infty$, $D_0 \to 0$. The ap-

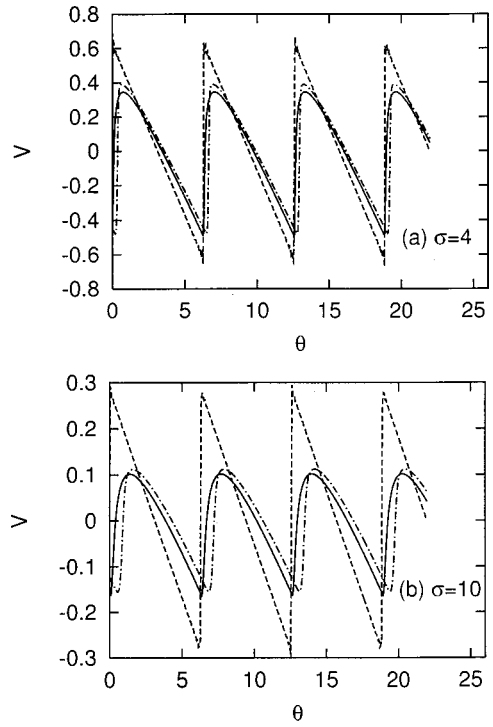


FIG. 7. Comparison of the postshock region approximate solution (solid lines), the Fay solution (dashed lines), and the numerical solution (dashed–dotted lines) in the time domain, $G_0=500$, $D_0=0.1$, (a) $\sigma=4$, (b) $\sigma=10$.

proximate solution is compared with the Fubini solution and the numerical one in the presented figures. Further, this paper contains the derivation of the new approximate solution of the generalized Burgers equation in the postshock region $\sigma>3$ for $D_0\ll 1$. This postshock approximate solution is contrasted with the Fay's solution and the numerical one in the presented figures as well. The postshock approximate solution provides more precise results than the Fay's solution in the limits of its validity because it takes into account the influence of the boundary layer which is not included in the Fay's solution. The postshock approximate solution converges to the Fay's solution for $D_0\to 0$. Both the presented approximate solutions extend the existing area in which it is known the approximate solution of the generalized Burgers equation.

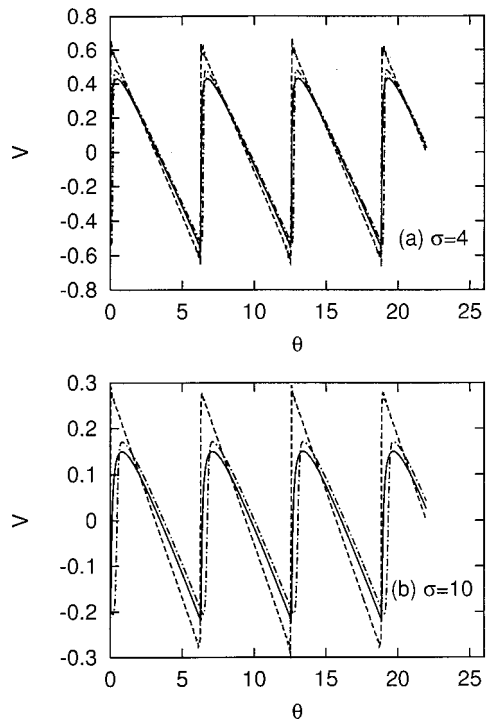


FIG. 8. Comparison of the postshock region approximate solution (solid lines), the Fay solution (dashed lines), and the numerical solution (dashed–dotted lines) in the time domain, $G_0=500$, $D_0=0.05$, (a) $\sigma=4$, (b) $\sigma=10$.

If we use for description of weakly nonlinear waves by the Khokhlkov–Zabolotskaya–Kuznetsov (KZK) model equation in tubes, we can see that the boundary layer causes diffraction effects as well. From this reason it is necessary to limit using of the GBE with respect to the cutoff frequency. On the basis of the KZK numerical solution analysis it is apparent that the transversal velocity component grows significantly when the cutoff frequency is exceeded. No longer we can assume in this case that propagating waves remain plane or quasiplane and hence the GBE cannot be used. Thus if the plane wave with the frequency below the cutoff one propagates in the small radius tube, where the influence of the boundary layer is dominant, then a distribution of acoustic velocity along the radius varies only a little and the results of the KZK and GBE are comparable.

ACKNOWLEDGMENTS

This work was supported by a grant of GACR No. 202/01/1372.

[1] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. (Pergamon, New York, 1987).
[2] O. V. Rudenko and S. I. Soluyan, *Theoretical Foundations of Nonlinear Acoustics* (Plenum, New York, 1977).
[3] K. Naugolnykh and L. Ostrovsky, *Nonlinear Wave Processes in Acoustics* (Cambridge University Press, Cambridge, 1998).
[4] M. F. Hamilton and D. T. Blackstock, *Nonlinear Acoustics* (Academic, New York, 1998).
[5] W. Chester, "Resonant oscillations in closed tubes," J. Fluid Mech. **18**, 44–64 (1964).
[6] D. T. Blackstock, *Nonlinear Acoustics (Theoretical)*, 3rd ed. (American Institute of Physics Handbook, McGraw Hill, New York, 1972).
[7] A. B. Coppens, "Theoretical study of finite-amplitude traveling waves in rigid-walled ducts: Behavior for strengths precluding shock formation," J. Acoust. Soc. Am. **49**, 306–318 (1971).
[8] D. A. Webster and D. T. Blackstock, "Finite-amplitude saturation of plane sound waves in air," J. Acoust. Soc. Am. **62**, 518–523 (1977).
[9] L. Gaete-Garreton and J. A. Gallego-Juarez, "Propagation of finite-amplitude ultrasonic waves in air. II. Plane waves in a tube," J. Acoust. Soc. Am. **73**, 768–773 (1983).
[10] S. H. Burns, "Finite-amplitude distortion in air at high acoustic pressures," J. Acoust. Soc. Am. **41**, 1157–1169 (1967).
[11] J. B. Keller and M. H. Millmann, "Finite-amplitude sound wave propagation in a waveguide," J. Acoust. Soc. Am. **49**, 329–333 (1971).
[12] J. B. Keller, "Nonlinear forced and free vibrations in acoustic waveguides," J. Acoust. Soc. Am. **55**, 524–527 (1974).
[13] A. H. Nayfeh and M. S. Tsai, "Non-linear wave propagation in acoustically lined circular ducts," J. Sound Vib. **36**, 77–89 (1971).
[14] A. H. Nayfeh and M. S. Tsai, "Nonlinear acoustic propagation in two-dimensional ducts," J. Acoust. Soc. Am. **55**, 1127–1133 (1974).
[15] J. H. Ginsberg, "Finite amplitude two-dimensional waves in a rectangular duct induced by arbitrary periodic excitation," J. Acoust. Soc. Am. **65**, 1127–1133 (1978).
[16] J. H. Ginsberg and H. C. Miao, "Finite amplitude distortion and dispersion of a nonplanar mode in a waveguide," J. Acoust. Soc. Am. **80**, 911–920 (1986).

[17] 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).
[18] M. F. Hamilton and J. A. TenCate, ''Finite amplitude sound near cutoff in higher-order modes of a rectangular duct,'' J. Acoust. Soc. Am. **84**, 327–334 (1988).
[19] A. H. Nayfeh, ''Nonlinear propagation of a wave packet in a hard-walled circular duct,'' J. Acoust. Soc. Am. **57**, 803–809 (1975).
[20] M. A. Foda, ''Analysis of nonlinear propagation and interactions of higher order modes in a circular waveguide,'' Acustica **84**, 66–77 (1998).
[21] D. T. Blackstock, ''Generalized Burgers equation for plane waves,'' J. Acoust. Soc. Am. **77**, 2050–2053 (1985).
[22] S. N. Makarov and E. Vatrushina, ''Effect of the acoustic boundary layer on a nonlinear quasiplane wave in a rigid-walled tube,'' J. Acoust. Soc. Am. **94**, 1076–1083 (1993).
[23] P. Konicek, M. Bednarik, and M. Cervenka, ''Finite-amplitude acoustic waves in a liquid-filled tube,'' in *Hydroacoustics* (Naval Academy, Gdynia, 2000), Vol. 3, pp. 85–88.
[24] M. Ochmann and S. Makarov, ''Representation of absorption of nonlinear waves by fractional derivatives,'' J. Acoust. Soc. Am. **94**, 3392–3399 (1993).
[25] S. Makarov and M. Ochmann, ''Nonlinear and thermoviscous phenomena in acoustics, Part II,'' Acustica **83**, 197–222 (1997).
[26] N. Sugimoto, ''Burgers equation with a fractional derivative: hereditary effects on nonlinear acoustic waves,'' J. Fluid Mech. **125**, 631–653 (1991).
[27] D. T. Blackstock, ''Thermoviscous attenuation of plane, periodic, finite-amplitude sound waves,'' J. Acoust. Soc. Am. **36**, 534–542 (1964).
[28] Ya. M. Zhileikin, T. M. Zhuravleva, and O. V. Rudenko, ''Nonlinear effects in the propagation of high-frequency sound waves in tubes,'' Phys. Acoust. **26**, 32–34 (1980).
[29] D. Gonghuan, ''Fourier series solution of Burger's equation for nonlinear acoustics in relaxing media,'' J. Acoust. Soc. Am. **77**, 924–927 (1985).
[30] F. H. Fenlon, ''A recursive procedure for computing the nonlinear spectral interactions of progressive finite-amplitude waves in nondispersive fluids,'' J. Acoust. Soc. Am. **50**, 1299–1312 (1971).

Existence of Mach cones and helical vortical structures around the underexpanded circular jet in the helical oscillation mode

Yoshikuni Umeda[a)]
Department of Aeronautics and Astronautics, Graduate School of Engineering, Kyoto University, Yoshida Honmachi, Sakyo-ku, Kyoto 606-8501, Japan

Ryuji Ishii
Department of Resources and Energy, Graduate School of Energy, Kyoto University, Yoshida Honmachi, Sakyo-ku, Kyoto 606-8501, Japan

(Received 23 February 2002; revised 22 April 2002; accepted 25 April 2002)

Existence of Mach cone and helical vortical structure in the helical oscillation mode of an underexpanded circular jet was confirmed by using schlieren instantaneous photographs and drawing of the envelopes of the Mach cones by the superposition of spherical sound waves radiated from two moving sound sources about the jet axis at a supersonic speed. Existence of such structures was conjectured in our earlier paper [Umeda and Ishii, J. Acoust. Soc. Am. **110**, 1845–1858 (2001)]. The envelopes of a Mach cone are observed as a V-shaped pattern composed of a pair of clear fine lines starting from a prominent point, which rotates about the jet axis. The helical vortical structure is observed as a bright pattern of the gathering of the tiny specks around the jet. It always appears to overlap on the envelopes of the moving Mach cones. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1487839]

PACS numbers: 43.28.Mw [MSH]

I. INTRODUCTION

The acoustic emission from an underexpanded free jet generally has a spectrum that contains very powerful discrete tones called *screech*. In a supersonic jet exhausting from a circular convergent nozzle, there are several shock cell structures; the large coherent vortical structures are convected downstream and interact with the oblique shock waves at the rear edge of the shock cells. Very strong acoustic waves are produced by the interaction between the vortical structures and the shock waves. The produced sound waves propagate upstream and stimulate the sensitive thin shear layer near the nozzle lip. In such a way, a feedback loop proposed by Powell (1953) is closed.

In the frequency characteristics of screech tones radiated from a circular underexpanded jet, Powell (1953) found several discontinuities. These discontinuities suggest that the circular jet oscillates in different oscillation modes in the different pressure ratio ranges. Experimental investigations about the instability of supersonic circular jets have been carried out by several researchers.

Davies and Oldfield (1962) found five stages exist in a pressure ratio range of less than about 6.0 and the first four oscillation modes are axisymmetrc, axisymmetric, lateral, and helical, respectively, with increasing pressure ratio. Powell *et al.* (1990, 1992) showed that the last oscillation mode at the highest pressure ratio R (dominant mode D) is lateral. They also showed that the plane of oscillation of the last lateral mode (mode D) rotates, vibrates, and dwells, while the first lateral mode (dominant mode B) occasionally changes to a helical mode at the same frequency.

For the circular jet, Shen and Tam (1998) carried out a numerical simulation on the generation of the screech tones for the axisymmetric mode and showed that their numerical results for the principal features of the screech tone such as sound intensity and frequency agree well with the experimental ones obtained by Ponton and Seiner (1992).

Westley and Wooley (1975, 1976) performed detailed study about the helical oscillation mode of a circular jet at a pressure ratio of $R=4.7$. For the axisymmetric and helical oscillation modes of screech tone, Panda (1998) investigated the periodic oscillation of shock waves in screeching underexpanded circular jets by using a shock detection technique based on laser light scattering by shock waves and showed that the mode of shock motion is the same as that of the emitted screech tone.

Panda (1999) also found from traversing microphone measuring and schlieren photography that partial interference between the upstream-propagating sound waves and downstream-propagating hydrodynamic waves manifests itself as a standing wave along the jet boundary. He derived an exact screech frequency formula in terms of the standing wavelength. Raman (1997, 1999) also found the same standing wave pattern along the jet boundary in the underexpanded rectangular jets as that found by Panda (1999) for circular jets.

Recently, we (Umeda and Ishii, 2001) investigated about the sound sources of the screech tones radiated from choked circular jets by using schlieren photography. From a series instantaneous schlieren photographs, we showed that the dominant sound source for the helical oscillation mode of the jet rotates about the jet axis in a plane just downstream of the rear edge of the third shock cell at a supersonic speed and the

a)Electronic mail: umeda@kuaero.kyoto-u.ac.jp

envelopes of the Mach cone moving with the sound source form helical-shaped wave fronts of the screech tone.

Furthermore, we (Umeda and Ishii, 2001) showed that for the dominant "*B*" mode tone, the axial distance from the nozzle exit to the location of the dominant sound source differs with each other depending on the oscillation mode at the same frequency (lateral or helical mode). For the helical oscillation mode, the axial distance is nearly equal to $4.2d$, which is equal to 1.5 times wavelength of the sound. For the lateral oscillation mode, it exists at about $2.8d$ from the nozzle exit, which is equal to one wavelength of the sound.

However, our previous paper (Umeda and Ishii, 2001), conjectured the existence of Mach cone and helical vortical structure around the underexpanded circular jet oscillating in the helical mode. In this paper, we tried to recognize the existence of such a Mach cone rotating around the jet and helical vortical structure moving downstream along the jet axis by the detailed analysis of the schlieren photographs taken at two angular positions of 90 and 70 degrees from the jet axis. Furthermore, the antisymmetricity of the directionality of the screech tone for the helical oscillation mode of the jet will be described in conjunction with the existence of the moving Mach cone.

II. EXPERIMENTAL APPARATUS

The convergent circular nozzle used in this experiment is the same as that used in our previous paper (Umeda and Ishii, 2001) and had an internal exit diameter of $d = 10$ mm. The external diameter in the exit plane was 26 mm. In this experiment, the cold air jet was exhausted from this nozzle at a pressure ratio $R = 3.90$. The photographs were taken using a conventional single-pass schlieren system and a digital camera Olympus type C-3030Zoom with about 3 mega piccels. This schlieren system used 20-cm-diam mirrors of 200-cm focal length, with a Sugawara type MS-230 spark light source of duration about 1 μs.

III. EXPERIMENTAL RESULTS

In our previous paper (Umeda and Ishii, 2001), we studied mainly about the dominant sound source of the screech tone in the helical oscillation mode by a series photographs and showed that the Mach cone rotates about the jet axis with a point sound source at a supersonic speed; a half angle of Mach cone was estimated to be about 30 degrees; and the vortical structure moves downstream along the jet axis. In the present paper, in order to recognize the existence of such Mach cones and helical vortical structures, the optical observation of the flow structures of the jet and the near sound field was performed by using instantaneous schlieren photographs.

Figure 1(a) shows an enlarged photograph of that shown in Fig. 2(a). This photograph shows the flow field of the supersonic jet issuing from a circular nozzle at a pressure ratio $R = 3.90$ and the near sound field of the screech tones radiated from this jet. This photograph was taken at a position perpendicular to the jet axis by using a Xe flash lamp with an exposure of about 1 μs. In this photograph, the flow structures such as the conical shock waves at the rear edge of the shock cells, the clear fine lines, and the bright and pale helical vortexlike structures around the jet are observed.

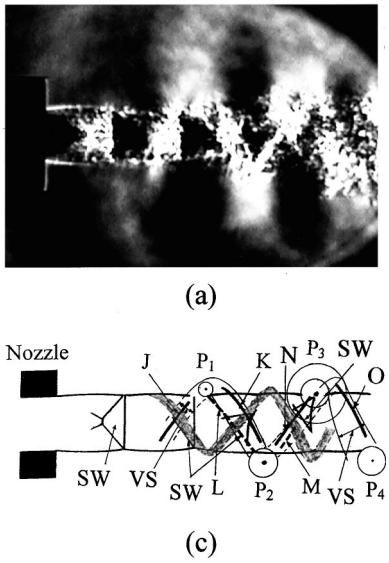


FIG. 1. Jet oscillating in the helical mode ($R = 3.90$) and its schematic view [enlarged photograph of Fig. 2(a) and schematic view of Fig. 2(c)].

These flow structures observed in the photograph of Fig. 1(a) are shown schematically in Fig. 1(c). In this figure, the bold lines with arrows and denoted by the symbols "*J*," "*K*," "*L*," "*M*," "*N*" and "*O*" indicate the clear fine lines. They construct three V-shaped patterns. The notations "P_1"–"P_4" indicate the prominent points. The parallel fine lines with arrows between two lines and denoted by the symbol "*VS*" show the helical vortexlike structure observed as the bright pattern of the gathering of many tiny specks. The pale vortexlike structure is shown by the parallel dotted lines with gray zone. The conical shock waves (reflected shock waves) observed at the rear edges of the shock cell structures are denoted by the symbol "*SW*." Further, the continuous lines indicate the flow structures observed in the front side of the jet and the dotted lines indicate those observed in the backside of the jet, respectively.

Figures 2(a) and (b) show the flow field of the super-

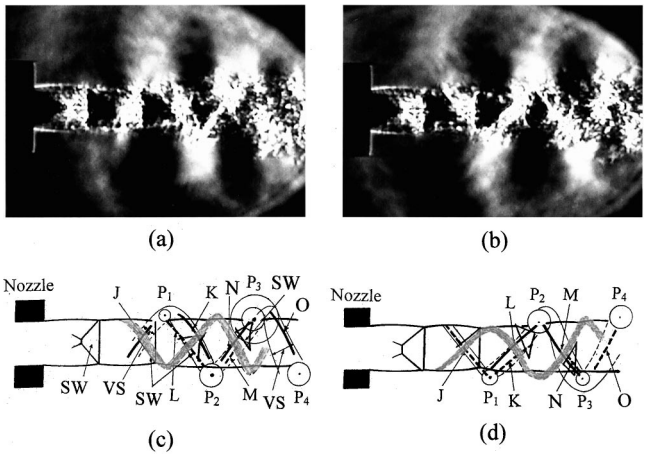


FIG. 2. Photographs of jets oscillating in the helical mode ($R = 3.90$) and their schematic views at different two instants separated by a half-cycle of the screech tone.

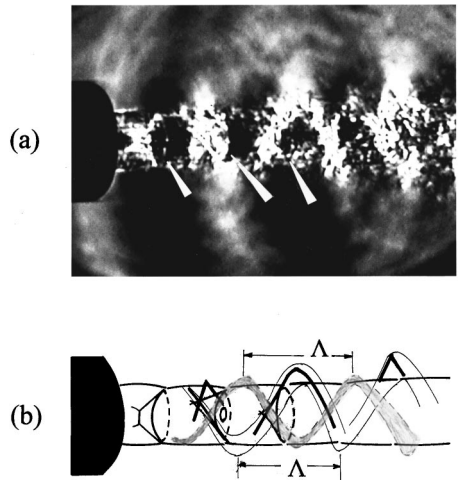


FIG. 3. Three-dimensional structure of the jet observed from an angular position of 70 degrees from the jet axis and its schematic view ($R=3.90$).

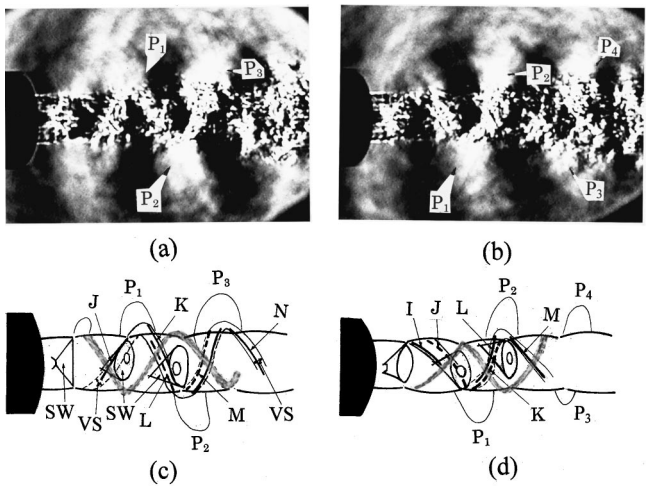


FIG. 4. Three-dimensional structure of the prominent points.

sonic jet issuing from a circular nozzle at a pressure ratio $R=3.90$ and the near sound field of the screech tones radiated from the jet at different two instants separated by a half-cycle of the screech tone. Figure 2(a) shows the same photograph as that of Fig. 1(a). Figures 2(c) and (d) show schematic views of the jets obtained from the photographs of Figs. 2(a) and (b), respectively.

In the photograph of Fig. 2(a), a bright M-shaped pattern of the vortexlike structure is observed to overlap along the clear fine lines [''J''–''O'' shown in Figs. 1(c) and 2(c)]. In contrast to this, a W-shaped pattern is observed in the photograph of Fig. 2(b) at a later time by about a half-cycle of screech tone from the instant shown in Fig. 2(a). Obviously, these patterns are axisymmetrical to each other.

In Fig. 2(a), a prominent point "P_2" is seen at just downstream of the rear edge of the third shock cell on the lower side the jet. In our previous paper (Umeda and Ishii, 2001), it was shown that this prominent point "P_2" is a dominant point sound source and it rotates about the jet axis in a plane perpendicular to the jet axis just downstream of the third shock cell. Two clear fine straight lines "L" and "M" making an upward V-shaped pattern are observed on both sides of the prominent point "P_2." The angle between these two lines is about 60 degrees. Two more downward V-shaped patterns constructing a part of an M-shaped pattern along with the former V-shaped pattern can be seen near the neighboring prominent points "P_1" and "P_3." The angle between the lines "J" and "K," and "N" and "O," making two V-shaped patterns is also about 60 degrees. Note that in the photographs of Figs. 2(a) and (b), there is a small space between two lines "L" and "K" constructing a part of two V-shaped patterns opposing to each other. In Fig. 2(a), it is observed that the prominent points "P_1" and "P_3" have donutlike pattern.

Figure 3(a) shows a photograph indicating three-dimensional structure of the circular jet taken at 70 degrees from the jet axis and Fig. 3(b) shows an illustration of the jet referring to the photograph of Fig. 3(a). The pressure ratio R of the jet is 3.90. In this photograph, three elliptic-shaped rear edges of the conical shock waves are observed around the jet axis (indicated by arrows). They are arranged at the same interval along the jet axis. Obviously, the actual shape of the rear edge of the conical shock wave (reflected shock wave) is a circle. The bright and pale vortical structures are observed along the jet axis. They give an appearance of double helix structure. These vortical structures have the same wavelength of Λ and they have a phase difference of about 180 degrees to each other.

Photographs of Figs. 4(a) and (b) show the three-dimensional structure of the jet taken at an angular position of 70 degrees from the jet axis. These two photographs correspond to those shown in Figs. 2(a) and (b), respectively, which were taken at an angular position of 90 degrees from the jet axis. Figures 4(c) and (d) show the schematic views of the jet observed in the photographs of Figs. 4(a) and (b), respectively. The notations and the lines used in the schematic views are the same as those used in Figs. 1 and 2.

From two photographs of Fig. 4 taken at 70 degrees from the jet axis, obviously the prominent points appear to be more active than those observed in the photographs taken at 90 degrees from the jet axis as shown in Figs. 1 and 2. They are not points, but they stretch along the clear fine lines around the jet and pile up largely from the jet boundary. So, they might be called as the prominent objects instead of the prominent points as described in our previous paper.

Figure 5 shows the photographs of a prominent point "P_0" appearing near the rear edge of the first shock cell and accompanied fine lines. These structures were not observed in our previous paper (Umeda and Ishii, 2001). Figure 5(a) shows an inclined photograph of the jet taken at an angular position of 70 degrees from the jet axis and Fig. 5(b) shows that taken at 90 degrees from the jet axis. Helical-shaped wave front of the screech tone is observed in the photograph of Fig. 6 (shown by an arrow).

IV. DISCUSSIONS

In the following, the existence of the moving Mach cones and helical vortical structures around the jet in the helical oscillation mode of the jet will be discussed.

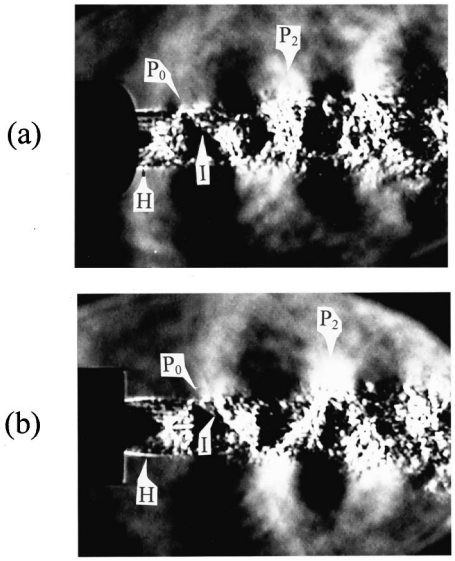


FIG. 5. Prominent point "P_0" and accompanied Mach cone.

A. Existence of the moving Mach cone

In the photographs of Figs. 2(a) and (b), clear fine lines ("*J*," "*K*," "*L*," "*M*," "*N*" and "*O*") making three V-shaped patterns are observed on the jet. These V-shaped patterns were recognized in the photographs shown in our previous paper (Umeda and Ishii, 2001), if we viewed them more carefully. As was described in Sec. III, the angle between these three pair of lines ("*J*" and "*K*," "*L*" and "*M*," "*N*" and "*O*") is about 60 degrees. Since a half-angle of the moving Mach cone was estimated as about 30 degrees in our previous paper (Umeda and Ishii, 2001), this value is very close to the estimated one. So, this fact suggests that these clear fine lines show the envelopes of the Mach cones rotating about the jet axis with moving point sound sources "P_1," "P_2" and "P_3."

If these three V-shaped patterns shown in Fig. 2(a) are envelopes of three moving Mach cones as was expected in our previous paper, these V-shaped patterns will reverse the direction to those shown in Fig. 2(a) after the point sound sources rotate about the jet axis by a half-cycle and M-shaped pattern will change to W-shaped pattern. This W-shaped pattern is surely observed in the photograph of Fig. 2(b). From the observation of these photographs, the moving Mach cones seem to surely exist. However, as the parallel fine lines "*L*" and "*K*" seen in the region between

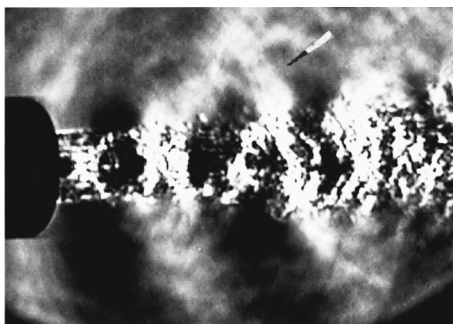


FIG. 6. Radiation of the helical-shaped wave front of the screech tone.

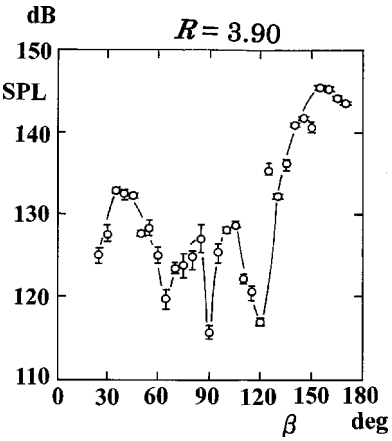


FIG. 7. Directionality of the screech tone in the helical mode [$R=3.90$, reproduced from Umeda and Ishii (2001)].

two point sound sources "P_1" and "P_2" are mysterious for us, we could not be sure of the existence of the moving Mach cone only from these photographs.

B. Antisymmetry of the directionality of the screech tone in the helical oscillation mode

The directionality of the screech tone in the helical oscillation mode was measured and the result was presented in our earlier paper (Umeda and Ishii, 2001) and is reproduced in Fig. 7 in this paper. As was described in the earlier paper, the idea of the single moving Mach cone can partly explain the directionality of the screech tone in the helical oscillation mode of the jet: maximum SPL at angles of 30 and 150 degrees; minimum SPL at angles of 60 and 120 degrees as shown in Fig. 7. But, it is difficult to understand the antisymmetry of the observed directionality of the screech tone in the helical "*C*" mode in the upstream and downstream directions by using the idea of the single moving sound source at a supersonic speed. Because, as was pointed out in our previous paper (Umeda and Ishii, 2001), the sound waves produced at the envelope of the single moving Mach cone must have the same intensity both in the upstream and downstream directions. Such situations are observed on the lower side of the jet shown in Fig. 8(b) and on the upper side of the jet shown in Fig. 8(e) of the series photographs reproduced from our previous paper.

In order to understand the antisymmetry of the directionality of the helical "*C*" mode tone, the series of schlieren photographs shown in Fig. 8 [reproduced from our previous paper (Umeda and Ishii, 2001)] must be investigated carefully: In Fig. 8(b), it is observed that two strong sound waves with symmetrical wave fronts (shown by arrows) are generated from the second prominent point "P_2" and are propagated in the upstream and downstream directions on the lower side of the jet. When the upstream-propagating sound wave passes through the rear edge of the second shock cell, as observed in Fig. 8(c), the original sound wave generated by the second prominent point "P_2" is intensified by the superposition with the sound wave produced by the first prominent point "P_1" appearing in the slightly downstream of the rear edge of the second shock cell. Such a situation is

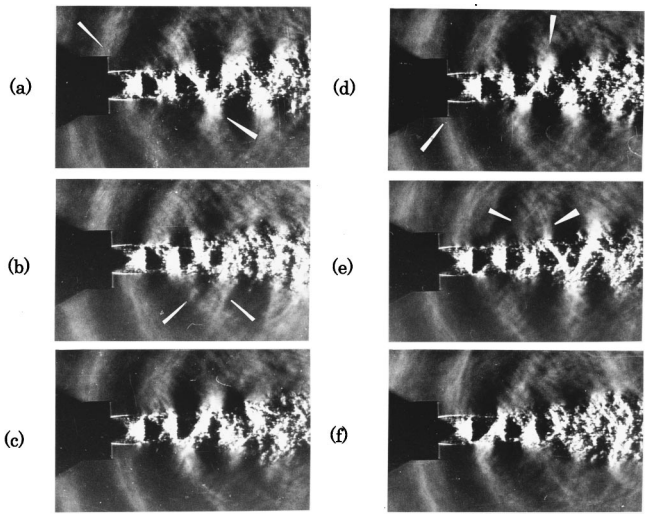


FIG. 8. One cycle of the screech tone generation in the helical oscillation mode for $R=3.90$ [reproduced from Umeda and Ishii (2001)].

also observed in Figs. 8(e) and (f). So, it can be understood that the screech tone in the helical oscillation mode is intensified in the upstream direction by the superposition of the sound waves generated by the second and first prominent points "P_2" and "P_1."

C. Drawing of the Mach cone by many spherical waves

In the following, we will try to show the antisymmetry of the directionality of the screech tone by the superposition of the spherical sound waves radiated from two moving sound sources "P_1" and "P_2" by drawing many circles (spherical waves) with different diameters.

First of all, a well-known Mach cone where a sound source moves along a straight line at a supersonic speed is shown in Fig. 9(a). In this figure, the centers of circles are arranged along a straight line at the same interval. The spherical wave fronts are drawn by circles with different diameters proportional to the passage of time from the generation of the point sound source. In this case, envelopes of the Mach cone are drawn by two bold straight lines connected with tangent lines of many circles with different diameters.

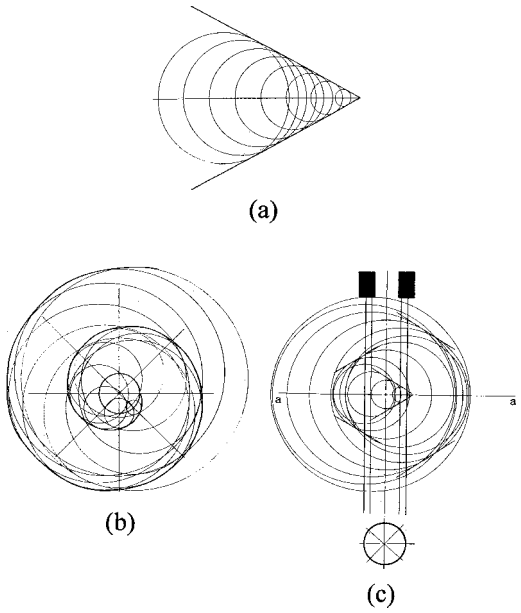


FIG. 9. Drawing of the Mach cone for a sound source moving along; (a) straight line; (b) circle (top view); (c) circle (side view).

Figure 9(b) shows a top view of the envelope of a Mach cone when a point sound source moves along a circular orbit at a supersonic speed. In this figure, a bold line circle in the center indicates a circular orbit of a moving sound source. The diameter of the circular orbit in this figure is proportional to the actual one observed in the photographs of Figs. 2–6 and 8. A sound source moves along this orbit at a supersonic speed, 715 m/s, as described in our previous paper (Umeda and Ishii, 2001).

In Fig. 9(b), the spherical sound waves are drawn by many circles with different diameters changing at every one out of every eight cycles of the screech tone. The position of the moving point sound source changes along the circular orbit at every one out of every eight cycles of the screech tone in the counterclockwise direction looked from the upstream of the jet. In this case, the envelope of these spherical waves or the Mach cone becomes a helix drawn by a bold line surrounding the circular orbit of the sound source.

Figure 9(c) shows a side view of the envelope of the single moving Mach cone at one instant. In this figure, two bold straight lines (vertical) show the outer limits of the circular orbit shown in the lower figure by a bold line circle. Fine straight lines including the central axis (vertical) show the radial positions of the sound source at every 1/8 cycles of the screech tone. A point sound source moves along the circular orbit in the plane indicated by a symbol "$a-a$." In this figure, the diameter of spherical waves increases proportionally with the passage of time from the generation of the point sound source and the envelope of the spherical waves is shown by several short lines. These short bold lines actually show some parts of the continuous envelope of the Mach cone rotating about the jet axis or the helical-shaped wave front of the screech tone propagating both in the upstream and downstream directions as shown in our previous paper (Umeda and Ishii, 2001). Therefore, actually these lines must be connected with each other both in the upstream and downstream regions. The intensity of the envelopes of the Mach cone drawn by these bold lines defined here as the density of the spherical waves is the same at the symmetric positions from the plane "$a-a$." So, this figure indicates that for the single moving sound source the directionality of the screech tone becomes a symmetrical one.

Figure 10 shows the time sequence of the envelope of the Mach cone in one cycle of the screech tone for the single moving sound source (side view). In this figure, time passes in alphabetical order. Figure 11 shows the relationship between the transfer distance "s" of an envelope of the Mach cone from the plane of "$a-a$" and the passage of time "t." The distance "s" was measured along the right-handed vertical bold lines as shown in Fig. 10(c). In Fig. 11, the transfer distance "s" shows the actual distance in mm and the time "t" represents the actual passage of time in μs and the notations "a"–"h" indicate the subnumbers in Fig. 10. From

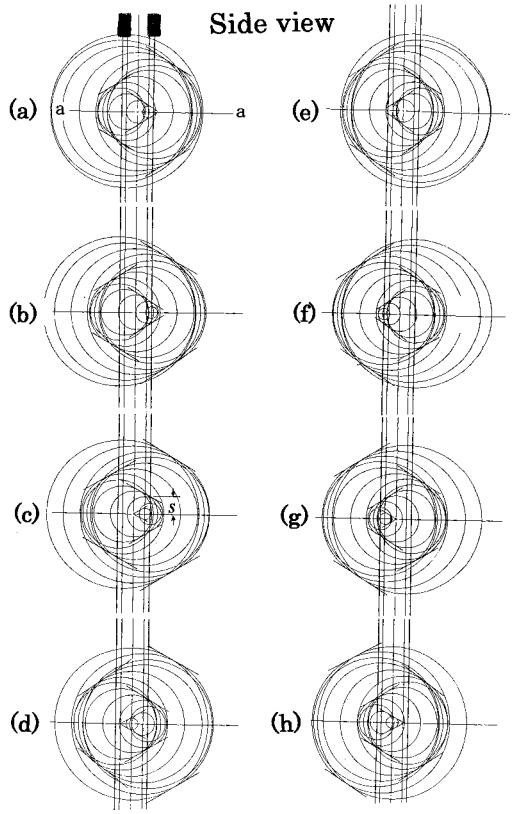


FIG. 10. Time sequence of the propagation of the Mach cone rotating about the jet axis (single sound source, side view).

this figure, it is observed that the envelope of the Mach cone moves at a constant speed, 412 m/s, the phase velocity of sound along the jet axis.

Now, let us consider the superposition of the sound waves radiated from two prominent points (sound sources) ''P_1'' and ''P_2'' moving along two circular orbits. As described in Sec. IV B, the antisymmetry of the measured directionality of the screech tone in the helical oscillation

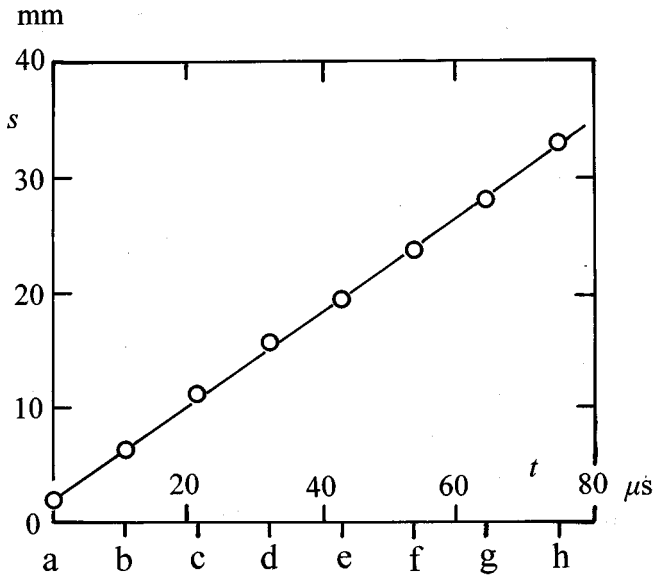


FIG. 11. Relationship between the transfer distance ''s'' of the Mach cone along the jet axis and time ''t.'' The radius of circular orbit is 0.96 cm.

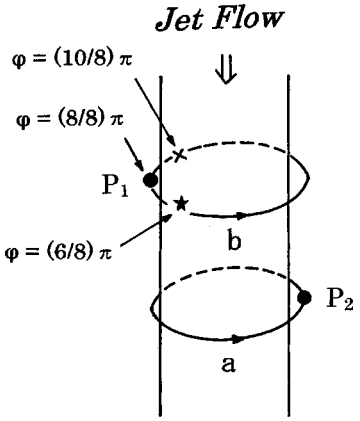


FIG. 12. Schematic view of two rotating sound sources about the jet axis.

mode can be explained by the superposition of the sounds radiated from mainly two moving sound sources ''P_1'' and ''P_2.'' The schematic view of the rotating two sound sources is shown in Fig. 12 referring to Fig. 8(c).

In this figure, two sound sources ''P_1'' and ''P_2'' rotate at the same speed ''v'' along the circular orbits ''b'' and ''a'' separated by an axial distance ''h.'' At the instant shown in Fig. 8(c), it is observed that the point source ''P_1'' appears about a half-cycle after the appearance of the point source ''P_2.'' So, the phase difference ''ϕ'' of the sound waves radiated from these sound sources is about 180 degrees $[=(8/8)\pi]$.

Figures 13(a) and (b) show the same photographs as those demonstrated in Figs. 2(a) and (b). Figures 13(c) and (d) show the drawings of the side view of the envelopes of the moving Mach cones created by two moving sound sources ''P_2'' and ''P_1'' at the instants corresponding to the above jets, respectively. These figures were drawn for the phase difference of $\phi=(\frac{6}{8})\pi$. In these drawings, it is observed that the wave fronts of the spherical waves crowd more densely in the upstream region than those observed in the downstream region. Namely, the envelopes of the Mach cones or the helical-shaped wave fronts of the screech tone become stronger in the upstream direction than that in the

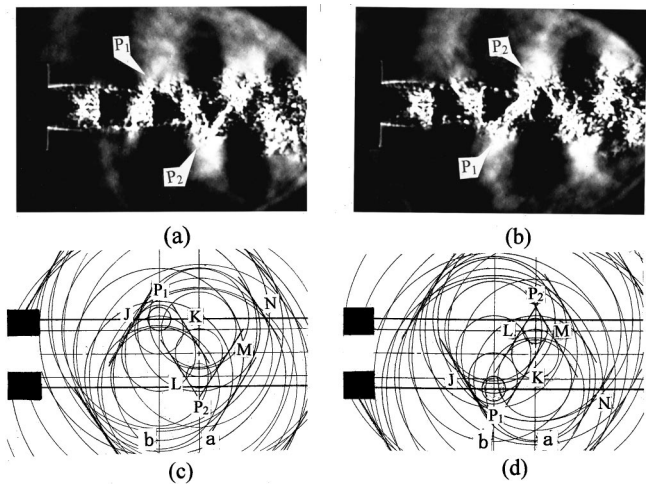


FIG. 13. Jets oscillating in the helical mode ($R=3.90$) and drawings of Mach cones for two rotating sound sources about the jet axis.

downstream direction as observed in the series photographs shown in Fig. 8.

Furthermore, it is observed that the patterns of the envelopes of the Mach cones shown in the drawings of Figs. 13(c) and (d) are very similar to those shown in the photographs of Figs. 13(a) and (b), respectively: Two prominent points "P_1" and "P_2" observed in Figs. 13(a) and (b) are seen at the corresponding positions shown in Figs. 13(c) and (d), respectively. Two V-shaped patterns constructing a part of M-shaped pattern shown in Fig. 13(a) and that of W-shaped pattern shown in Fig. 13(b) are observed at the corresponding positions shown in Figs. 13(c) and (d), respectively. Also, the mysterious double structure [parallel lines "*K*" and "*L*" shown in Figs. 2(c) and (d)] of the clear fine lines observed in the photographs of Figs. 2(a) and (b) or Figs. 13(a) and (b) are seen in the region between two prominent points "P_1" and "P_2" in the drawings of Figs. 13(c) and (d), respectively.

From the very good agreement between the flow patterns observed as the clear fine lines in the photographs and the envelopes of the moving Mach cones seen in the corresponding drawings obtained by the superposition of the spherical sound waves radiated from two moving point sound sources "P_1" and "P_2," the drawings for the envelopes of the Mach cones simulate very well the clear fine lines observed in the photographs of Figs. 2(a) and (b) or Figs. 13(a) and (b).

Therefore, it is confirmed that two V-shaped patterns composed of the clear fine lines "*J*," "*K*," "*L*" and "*M*" found in photographs of Figs. 2 and 13 are surely envelopes of the Mach cones rotating about the jet axis at a supersonic speed. It is also confirmed that although five prominent points are observed vividly in this experiment, the antisymmetry of the observed directionality of the screech tone shown in Fig. 7 can be explained by the superposition of the sound waves mainly radiated from two point sound sources "P_1" and "P_2."

Now, it is easily understood that the donutlike patterns observed in the photographs of Figs. 2(b) and 13(b) indicate the back view of the moving Mach cones with the prominent points "P_1" and "P_3." So, in this case, these Mach cones rotate about the jet axis in the counterclockwise direction looked from the upstream of the jet. Since these Mach cones are generated by the rotation of the sound sources about the jet axis, the envelopes of the Mach cones must be bent around the jet boundary. Such a situation is clearly seen in the inclined photographs of Figs. 3 and 4.

D. Existence of one more point sound source near the rear edge of the first shock cell

So far, in our previous paper (Umeda and Ishii, 2001) as well as in the present paper, five prominent points ("P_1"–"P_5") have been observed in the schlieren photographs. They appear slightly downstream of the rear edges of the second, third, fourth, fifth, and sixth shock cells, respectively. Why does the prominent point not appear near the rear edge of the first shock cell? This is a quite unnatural phenomenon. So, many schlieren photographs were examined again carefully to find the prominent point and the accompanied envelopes of the Mach cone near the rear edge of the first shock cell. Figure 5 just shows the photograph showing such a prominent point and the envelopes of the Mach cone. Figure 5(a) shows an inclined photograph of the jet taken at an angular position of 70 degrees from the jet axis and Fig. 5(b) shows a photograph of the jet taken at 90 degrees from the jet axis. In both photographs, the prominent point denoted by the symbol "P_0" is observed slightly downstream of the rear edge of the first shock cell as was expected. The accompanied envelopes of the Mach cone can also be seen as two pale straight lines denoted by the symbols "*H*" and "*I*." The angle between these two lines is also about 60 degrees.

So, two envelopes of this Mach cone rotate about the jet axis with the rotating prominent point "P_0" and one of two envelopes of the Mach cone seen in the upstream of the prominent point "P_0" sweeps the sensitive thin shear layer near the nozzle lip and the nascent helical-shaped vortex will be generated by the stimulation of the rotating Mach cone accompanied with the point sound source "P_0." This generation mechanism of the coherent vortical structure is different from that in the usual feedback loop proposed by Powell (1953). In the feedback loop proposed by Powell (1953), when the upstream-propagating sound waves pass through the nozzle lip, the sensitive thin shear layer near the nozzle lip is stimulated by the sound waves and the coherent vortical structure is generated.

Now, we will propose another generation mechanism of the coherent vortical structure as described in the above. One of two envelopes of the Mach cone "*H*" accompanied with the point sound source "P_0" can generate the coherent vortical structure by sweeping the sensitive thin shear layer near the nozzle lip. Once the dominant helical "*C*" mode oscillation of jet occurs, such a vortical structure can be created continuously by the sweep of the Mach cone rotating with the sound source "P_0." The generated coherent vortical structure is grown up partly by the excitation of the individual rotating Mach cones arranged alternately along the jet axis one after another as shown in the photographs of Fig. 4.

In the case of the axisymmetric and lateral oscillation modes of jets, the rotating Mach cone about the jet axis never exists. So, in these cases only stimulation by the upstream-propagating sound wave in the ambient can produce the coherent vortical structure. However, in the helical oscillation mode of the jet, there are two possibilities to generate the vortical structure: stimulation by

(1) the upstream-propagating sound waves in the ambient as was proposed by Powell (1953) and
(2) the sweep of an envelope of the Mach cone rotating with the point sound source "P_0."

Which mechanism is dominant to generate the vortical structure in the helical oscillation mode of jet? Although in the photograph of Fig. 5 a tiny bright spot is observed in the jet boundary and at the tip of the envelope of the Mach cone seen in the upstream of the point sound source "P_0," we cannot decide which mechanism is dominant from the photograph. At least for the present, we suppose that both

mechanisms join forces to generate the coherent helical-shaped vortical structure.

E. Existence of the helical vortical structure

In the above section, the existence of the moving Mach cone about the jet axis is confirmed. In the following, we will try to show the existence of the helical vortical structure.

In the photographs of Figs. 1–5 and 13, two types of vortical structures are observed. The bright vortexlike structure is observed to overlap on the envelopes of the rotating Mach cones with the point sound sources around the jet. In the photograph of Fig. 3 taken at an angular position of 70 degrees from the jet axis, these two types of the vortical structures can be seen clearly. They give an appearance of double helix structure like DNA (deoxyribonucleic acid). One is bright and the other is pale. The double structure of the vortical structures was not expected by the well-known feedback loop theory. Why does such a double structure of the vortices appear? Unfortunately, for the present, this mysterious phenomenon cannot be explained.

In the following, the bright vortical structure will be considered. At an instant shown in the photographs of Figs. 2(a) and 4(a) and the schematic views of the jet shown in Figs. 2(c) and 4(c), the oscillating conical shock wave "SW" at the rear edge of the third shock cell and the trough of the bright vortical structure are observed very close to each other on the lower side of the jet. At this instant, a very strong point sound source "P_2" also appears near both these structures in the lower side of the jet. At an instant of a half-cycle of the screech tone later, as shown in the photographs of Figs. 2(b) and 4(b) and the schematic views of the jet shown in Figs. 2(d) and 4(d), the conical shock wave and the crest of the bright vortical structure are seen again very close to each other on the upper side of the jet. At this instant, the point sound source "P_2" appears near both the structures on the upper side of the jet.

These situations indicate that the very strong point sound source "P_2" is produced by the interaction between the oscillating conical shock wave "SW" in the jet and the bright coherent vortical structure around the jet. This interaction occurs at every instant along the circular orbit in the plane perpendicular to the jet axis and slightly downstream of the third shock cell. All the point sound sources "P_0"–"P_5" are generated by the same mechanism, namely by the interaction between the individual oscillating conical shock waves and the bright coherent vortical structure. These point sound sources are observed to stretch along the clear fine lines around the jet and pile up largely from the jet boundary as shown in the photographs of Fig. 4. This fact shows that the interaction between the downstream-convecting bright coherent vortical structure and the oscillating conical shock wave does not occur at one point, but it occurs along a part of the circumference of the rear edge of the conical shock wave. The sound sources "P_0"–"P_5" so generated rotate about the jet axis in the individual planes perpendicular to the jet axis. These planes exist slightly downstream of the rear edge of the first to sixth shock cell structures. All the point sound sources rotate about the jet axis at a supersonic speed. So, the Mach cones are generated behind the individual moving sound sources as observed in the photographs of Figs. 1–5, 8, and 13.

Why do the bright vortical structures always tend to overlap on the envelopes of the rotating Mach cones? In the photographs of Figs. 2 and 13 taken at an angular position of 90 degrees from the jet axis, it is observed that the vortical structure convects downstream along the jet axis by 2 shock-cell lengths during one turn of the sound source about the jet axis. This means that the wavelength Λ of the vortical structure is just equal to 2 shock-cell lengths:

$$\Lambda = 2 \ \text{shock-cell lengths}. \tag{1}$$

In this case, six point sound sources "P_0"–"P_5" are observed to rotate about the jet axis in the individual planes perpendicular to the jet axis separated by 1 shock-cell length. They appear on both sides of the jet alternately along the jet axis. On one side of the jet, point sound sources are observed at every 2 shock-cell lengths, which is equal to the wavelength Λ of the vortical structure. These point sound sources rotate about the jet axis at a supersonic speed in the same direction. So, the adjacent envelopes of the moving Mach cones accompanied with the individual sound sources face the opposite directions and are connected with each other surrounding the jet boundary. In such a way, a cylindrical system composed of six rotating Mach cones is created around the jet. In this case, the system has a helical structure because each individual envelope of the rotating Mach cone forms the helical-shaped wave front of the screech tone as shown in our previous paper (Umeda and Ishii, 2001).

As described in the above, both the helical-shaped vortical structure and the helical-shaped Mach cone system have the same wavelengths of 2 shock-cell lengths. So, if a helical-shaped vortical structure convects downstream along the jet axis to pass the above helical-shaped Mach cone system, the vortical structure will be excited effectively by the rotating Mach cone system and the vortical structure will be grown up rapidly. In this way, the bright vortical structures are created along the individual envelopes of the rotating Mach cone system as shown in the photographs of Figs. 1–5, 8, and 13. Therefore, it is confirmed that the vortical structures appearing in the helical oscillation mode of the jet have surely helical structure. This bright vortical structure changes the shape with the rotation of the Mach cones about the jet axis as shown in Figs. 2 and 13.

As shown in Sec. IV B, the upstream-propagating sound wave radiated from the moving point sound source "P_2" positively interacts with the sound wave radiated from the moving point source "P_1" at slightly downstream of the rear edge of the second shock cell.

The intensified sound wave in such a process propagates upstream further and reaches at the nozzle lip. Then, a nascent vortical structure is generated by the stimulation of the upstream-propagating sound waves (or/and by the excitation of the sweep of the envelope of the moving Mach cone as described in the previous section), convects downstream, and grows rapidly by the excitation of the envelopes of the moving Mach cones. This grown-up bright vortical structure has the wavelength of Λ (2 shock-cell lengths) and interacts with

the oscillating conical shock waves in the jet to generate the point sound sources (say, "P_2" and "P_4") at every 2 shock-cell lengths (the wavelength Λ of the vortical structure) from the nozzle exit on one side of the jet as shown in Figs. 2(a) and (c).

At a half-cycle of the screech tone later, this bright vortical structure convects downstream by 1 shock-cell length (a half-wavelength of the vortical structure) and interacts with the oscillating conical shock waves to generate the point sound sources (say, "P_1" and "P_3") at every 2 shock-cell lengths (1 wavelength Λ of the vortical structure) from the nozzle exit on the other side of the jet as shown in Figs. 2(b) and (d). In such a way, the point sound sources are generated alternately on both sides of the jet along the jet axis. Therefore, it can be understood that the point sound sources appear on both sides of the jet alternately along the jet axis.

As described in the above, although the interaction between the downstream-convecting vortical structure and the oscillating conical shock waves occurs at every instant in the individual plane where one sound source rotates about the jet axis, such an interaction can clearly be seen at the instants when the interaction occurs on both sides of the jet as shown in the photographs of Figs. 2(a) and (b) or Figs. 13(a) and (b). Obviously, these photographs show that the interaction between the helical vortical structure with the wavelength Λ (2 shock-cell lengths) and the synchronously oscillating conical shock waves creates point sound sources with the spacing of 2 shock-cell lengths on each side of the jet.

Incidentally, from the photographs of Figs. 2–5, 8, and 13, it is observed that the excited vortical structure disintegrates rapidly in a short time, less than about a half-period of the screech tone. In this case, as the frequency of the screech tone was measured as 11.75 kHz, a half-period of the screech tone is about 0.043 ms. If the vortical structure does not disintegrate rapidly, the bright vortical structure ought to surround the jet completely so as to be looked through the jet.

V. CONCLUSIONS

The existence of the rotating Mach cone about the jet axis and the downstream-convecting helical-shaped vortical structure was conjectured in our previous paper (Umeda and Ishii, 2001). In the present paper, the existence of such structures was investigated by the instantaneous schlieren photographs. The following results are obtained in the present paper:

(1) The existence of the rotating Mach cone about the jet axis is confirmed by comparing instantaneous schlieren photographs with the drawings of the envelopes of the Mach cones obtained by the superposition of the spherical sound waves radiated from two moving sound sources "P_1" and "P_2" at a supersonic speed. The envelopes of the Mach cones accompanied with the point sound sources "P_0"–"P_5" are observed both in the previous and the present experiments.

(2) The point sound source "P_0" observed near the rear edge of the first shock cell structure gives the possibility of another generation mechanism of the nascent coherent vortical structure. The sweep of one of the two envelopes extending from the point sound sources "P_0" can stimulate the sensitive thin shear layer of the jet boundary near the nozzle lip and can generate the nascent coherent vortical structure. This generation mechanism of the vortical structure is applicable only to the helical oscillation mode of jet.

(3) Two types of the vortical structures are observed as bright and pale patterns composed of the gathering of many tiny specks around the jet. The bright vortical structure is observed along the envelopes of the moving Mach cones. This bright vortical structure has helical structure. The interaction between the downstream-convecting bright helical vortical structure and the oscillating conical shock waves at the rear edges of the shock cell structures generates at least six point sound sources in the planes slightly downstream of each shock cell. These point sources appear alternately on both sides of the jet along the jet axis and they rotate about the jet axis at a supersonic speed. Six Mach cones are generated behind six moving point sources and they form the helical-shaped wave fronts of the screech tone in the helical oscillation mode.

(4) Two point sound sources "P_1" and "P_2" are dominant ones. The superposition of the sound waves radiated from these moving sound sources at a supersonic speed causes mainly the antisymmetry of the directionality of the screech tone in the helical oscillation mode.

ACKNOWLEDGMENTS

The authors would like to thank Professor A. Powell at Houston University for valuable discussions and comments on this paper.

Davies, M. G., and Oldfield, D. E. S. (**1962**). "Tones from a choked axisymmetric jet. II. The self-excited loop and mode of oscillation," Acustica **12**, 267–277.

Panda, J. (**1998**). "Shock oscillation in underexpanded screeching jets," J. Fluid Mech. **363**, 173–198.

Panda, J. (**1999**). "An experimental investigation of screech noise generation," J. Fluid Mech. **378**, 71–96.

Powell, A. (**1953**). "On the mechanism of choked jet noise," Proc. Phys. Soc. London, Sec. B **66**, 1039–1056.

Powell, A., Umeda, Y., and Ishii, R. (**1990**). "The screech of round choked jets, revisited," AIAA Pap. 90–3980.

Powell, A., Umeda, Y., and Ishii, R. (**1992**). "Observation of the oscillation modes of circular jets," J. Acoust. Soc. Am. **92**, 2823–2836.

Raman, G. (**1999**). "Supersonic jet screech: Half-century from Powell to the present," J. Sound Vib. **225**, 543–572.

Shen, H., and Tam, C. K. W. (**1998**). "Numerical simulation of the generation of axisymmetric mode jet screech tones," AIAA J. **36**, 1801–1807.

Umeda, Y., and Ishii, R. (**2001**). "On the sound sources of screech tones radiated from choked circular jets," J. Acoust. Soc. Am. **110**, 1845–1858.

Westley, R., and Raman, G. (**1997**). "Cessation of screech in underexpanded jets," J. Fluid Mech. **336**, 69–90.

Westley, R., and Wooley, J. H. (**1975**). "The near-field sound pressures of a choked jet when operating in the spinning mode," AIAA Pap. 75–479.

Westley, R., and Wooley, J. H. (**1976**). "Sound pressures of a choked jet oscillating in the spinning mode," Astronaut. Aeronaut. **43**, 185–202.

Time-resolved tracking of a sound scatterer in a complex flow: Nonstationary signal analysis and applications

Nicolas Mordant and Jean-François Pinton[a)]
École Normale Supérieure de Lyon & CNRS UMR 5672, Laboratoire de Physique, 46 allée d'Italie, F-69364 Lyon, France

Olivier Michel
Laboratoire d'astrophysique & CNRS UMR 6525, Université de Nice, Parc Valrose, F-06108 Nice, France

(Received 1 November 2000; revised 4 December 2001; accepted 4 February 2002)

It is known that ultrasound techniques yield nonintrusive measurements of hydrodynamic flows. For example, the study of the echoes produced by a large number of particles insonified by pulsed wavetrains has led to a now-standard velocimetry device. In this paper, a new technique for the measurement of the velocity of individual solid particles moving in fluid flows is proposed. It relies on the ability to resolve in time the Doppler shift of the sound scattered by the continuously insonified particle. For this signal-processing problem two classes of approaches can be used: time-frequency analysis and parametric high-resolution methods. In the first class the spectrogram and reassigned spectrogram is considered, and applied to detect the motion of a small bead settling in a fluid at rest. In nonstationary flows, methods in the second class are more robust. An approximated maximum likelihood (AML) technique has been adapted, coupled with a generalized Kalman filter. This method allows for the estimation of rapidly varying frequencies; the parametric nature of the algorithm also provides an estimate of the variance of the identified frequency parameters. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1477932]

PACS numbers: 43.30.Es, 43.60.−c, 47.80.+v, 43.60.Qv [DLB]

I. INTRODUCTION

Our original motivation stems from the need of a technique to measure the velocity of individual tracer particles in complex flows, and thus to be able to study hydrodynamics in the Lagrangian frame of reference. Indeed, the Lagrangian dynamics governs the physics of transport, mixing, and the Eulerian complexity of chaotic and turbulent flows. Lagrangian studies have been made in numerical experiments where chaotic[1] and turbulent[2–5] flows have been considered. For turbulence, the numerical studies are limited to small Reynolds number flows. Although experimental measurements are much needed, few Lagrangian methods have been developed. The main ones rely on optical techniques: tracking features have been adapted to particle image velocimetry measurements[6] and fast detectors have been used to image short particle tracks and measure particle accelerations.[7] In the first case the time resolution is low (60 frames per second) and in the second case the volume in space sampled is small ($\sim$1 mm^3). Inspired by sonar principles, we propose an acoustic method that can resolve single-particle motion over extended regions of a space ($\sim$1000 cm^3) with a high time resolution (1 ms). The main advantages of our technique are: (i) the particle velocity is measured directly and not as a derivative of position (always quite noisy); (ii) the acoustic detection samples a larger region of space than optical devices and can record particle motion for longer times than high-speed cameras; (iii) it can be used in nontransparent media. The principle is to monitor the Doppler shift of the sound scattered by a particle which is *continuously* insonified. This is an extension of the pulsed Doppler principle that has been developed to measure velocity profiles and that has many applications in fluid mechanics and medicine.[8] The continuous insonification improves the time resolution of the measurement, although it limits the tracking to a very small number of particles.

The measurement relies on the ability to track a Doppler frequency and its variation in time. For this signal-processing problem two classes of approaches have been developed: (i) time-frequency analysis and (ii) high-resolution parametric spectral analysis. Time-frequency methods rely mainly upon the quadratic Wigner–Ville transform, or smoothed versions of it. Numerous studies and papers have recently been published, in which the theoretical issues are presented (see, e.g., the textbooks by Flandrin[9] or Cohen[10]). These nonparametric techniques are convenient and well-suited for weakly nonstationary signals with a good signal-to-noise ratio (SNR). However, time-frequency representations present numerous drawbacks when it comes to extracting trajectory information. Their quadratic nature gives rise to numerous spurious interference terms that require postprocessing. For signals with a faster frequency modulation and a low SNR, we show here that an optimized parametric approach is a better choice. Parametric high-resolution spectral analysis methods take advantage of an *a priori* knowledge of the spectral content of the recorded signal, namely the emitted signal frequency plus one or many Doppler-shifted echoes. Furthermore, a time-recursive frame for the estimation of the Doppler shift is proposed here, where the evolution of the frequency is taken into account in the algorithm.

The two methods are tested in two experiments, in which the acoustic signals have different time scales and

a)Electronic mail: pinton@ens-lyon.fr

 0001-4966/2002/112(1)/108/11/$19.00

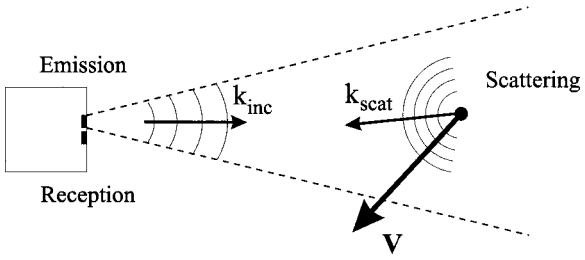


FIG. 1. Principle of measurement. A large 3D measurement zone is achieved by using a transducer size of a few wavelengths.

noise levels. The first experiment is a study of the transient acceleration of a heavy sphere settling under gravity in a fluid at rest. This situation addresses the fundamental problem of the response of a solid particle to a step force. The characteristic time scale of velocity variations is slow ($\tau \sim 50$ ms) and the signal-to-noise ratio is fair (about 20 dB); we show that a technique of reassignment of the spectrogram gives good results. The second experiment deals with the motion of a small, neutrally buoyant sphere embedded in a turbulent flow. In this case the fundamental issue is that of the Lagrangian velocity fluctuations. There, velocity variations occur over times of about 1 ms and the signal-to-noise ratio is low (less than 6 dB). We show that the approximate maximal likelihood (AML) parametric method yields very good results in that situation. Altogether, our techniques yields a direct measurement of the velocity of single particles in complex fluid flows. It allows investigation of experimentally fundamental issue in fluid dynamics (the study of flow motion in the Lagrangian frame) and has potential applications in fields such as engineering (transport and mixing) and biomedicine (the tracking of implants).

The paper is organized as follows: in Sec. II we present the acoustic technique and measurement procedure. In Sec. III we describe the signal-processing techniques, with a particular emphasis on the approximate maximum likelihood (AML) method, which has been developed and optimized to this particle tracking problem. Examples of applications to measurements in real flows are given in Sec. IV.

II. ACOUSTICAL SETUP

A. Principle of the measurement

In the experimental technique proposed here, a particle is continuously insonified. It scatters a sound wave whose frequency is shifted from the incoming sound frequency due to the Doppler effect. This Doppler shift is directly related to the particle velocity $\mathbf{v}_p$

$$\Delta\omega = \mathbf{q}\cdot\mathbf{v}_p, \tag{1}$$

where $\mathbf{q}$ is the scattering wave vector (the difference between the incident and scattered wave vectors $\mathbf{q}=\mathbf{k}_{scat}-\mathbf{k}_{inc}$) and ω is the wave pulsation.

We choose a backscattering geometry (see Fig. 1) so that $\mathbf{q}=-2\mathbf{k}_{inc}$ and the frequency shift becomes

$$\Delta\omega(t) = -2\frac{v(t)}{c}\omega_0, \tag{2}$$

where c is the speed of sound, ω_0 is the incident pulsation, and $v(t)$ is the component of the velocity on the incident direction at time t. We continuously insonify the moving particle and record the scattered sound. If need be, the particle position can be obtained by numerical integration of the velocity signal.

B. Transducer characteristics and acquisition

The transducers are made of two piezoelectric elements of size 2×2 mm each, separated by 100 μm. Their resonant frequency is about 3.2 MHz and their bandwidth at -3 dB is 1.5 MHz. Sound emission is set at 3 or 3.5 MHz; experiments are performed in water so that the wavelength is $\lambda=0.50$ or 0.43 mm. The corresponding emission cone for each $d=2$-mm-square element is 29° at 3 MHZ and 24° at 3.5 MHz. In our measurements, the particle to transducer distance lies between 5 and 40 cm, so that measurements are made in the far field ($d^2/\lambda > 10$ mm). Given maximum flow and particle velocities of the order of 1.5 $\mathrm{m\cdot s^{-1}}$, we expect a maximum sound-frequency shift of the order of 5 or 6 kHz, depending on whether the emission is at 3 or 3.5 MHz. This yields a frequency modulation rate of at most 0.25%. One transducer is used for continuous sound emission and the other for scattered sound detection. As the operation is continuous (as opposed to pulsed) and the elements are located close to one another, we observe a coupling between the emitter and the receiver of the order of 60 dB (this is due both to electromagnetic coupling and to acoustic surface waves cross talk).

When connected to a 50-Ω impedance, the acoustic transducers yield an electrical signal of about 8 mV. This signal is mainly due to the electromagnetic coupling with the emitter. The part of the signal due to the acoustic scattering from the moving particle is of the order of 2 to 30 μV—for comparison, the noise is 1 μV. Hence, the signal-to-noise ratio is between 0 and 30 dB. The transducer output is sampled at 10 MHz over a 21-bit dynamical range (input range 31.25 mV) and numerically heterodyned at the emitting frequency. It is then decimated at the final sampling frequency of 19 531 Hz. The acquisition device is an Agilent hpe1430A VXI digitizer.

C. Scattering by an elastic sphere

The study of sound scattered by a fixed solid sphere is a classic but continuing area of study, and difficulties arise in the interpretation of observed phenomena especially when trying to deal with elasticity and absorption.[11–13] Complex behavior is observed linked with resonances of Rayleigh waves at the surface of the sphere. As a consequence, the scattered pressure distribution varies both in directivity and amplitude. A generic expression for the far-field pressure is the following:

$$p_{scat}(r,\theta) = p_{inc}\frac{af(ka,\theta)}{2r}e^{ikr}, \tag{3}$$

where r is the distance from the center of the sphere, a its radius, p_{inc} the incident pressure on the sphere, k the incident wave number in the fluid, θ the scattering angle, and f is a

form function which depends on the physical properties of the solid medium. Under very general assumptions, f can be expressed analytically.[11] Physically, f represents the sum of the specular echo and of interferences due to the radiation by Rayleigh waves.[12,13] As a result, f is a strongly varying function, particularly for high values of ka. In our experiments we used spheres of different material (polypropylene PP, steel, tungsten carbide, glass) with corresponding ka between 7 and 15. The flow acts on the sphere motion, thus causing its acceleration and, eventually, its rotation. These effects may change the radiation diagram: first, there is Doppler shift for the sound received by the sphere, and, perhaps more importantly, the sphere rotation may change the Rayleigh emission. For these reasons, the evolution of the amplitude of the scattered sound during the particle motion is quite complex. However, the observed amplitude modulation varies slowly enough to allow a correct estimate of the frequency modulation of the scattered sound (see the times series presented in Sec. IV).

III. SIGNAL PROCESSING

Numerous spectral estimation techniques are based on the ideas behind Fourier analysis of linear time-invariant (LTI) differential equations. These techniques may be divided into (i) nonparametric techniques where the basis functions are implicitly the harmonically related complex exponentials of Fourier analysis, and (ii) parametric techniques whose task is the estimation of the parameters of a (sub)set of complex exponentials. The spectrogram and the reassigned spectrogram belong to the former category, whereas the maximum likelihood and its approximate form belong to the latter.

A. Time-frequency analysis

The most common time-frequency distribution (TFD), the spectrogram, involves a moving time window. This window attempts to capture a portion of the signal which is sufficiently restricted in time so that stationarity and LTI assumptions are approximately met. To overcome the inherently poor localization in the time-frequency plane, a method has been proposed by Gendrin *et al.*,[14] and extended more recently by Auger and Flandrin.[9,15] The idea is to locally reassign the energy distribution to the local center of gravity of the Fourier transform. Despite its ability to exhibit clear and well-localized trajectories in the time-frequency plane, this technique requires an additional image-processing step to extract the TF trajectory. For rapidly fluctuating frequency modulations and/or low SNR, spurious clusters appear which makes this extraction difficult. The parametric method presented below is more robust.

B. Parametric spectral estimation based on approximate likelihood

This approach is based upon maximum likelihood spectral estimation (see, e.g., Kay[16]). The fundamentals are briefly recalled, as they serve as a basis for the approximate likelihood scheme, originally developed by Clergeot and Tressens.[17] This work is extended here within a recursive estimation frame, thus allowing us to track the variations of the Doppler frequency shift induced by fast velocity changes of a scattering sphere imbedded in a turbulent flow. Michel and Clergeot have developed a similar approach for nonstationary spectral analysis in an array processing frame.[18,19]

1. Introduction

In this section, we address the problem of estimating an unknown number M of frequencies (unknown parameters in the sequel) $f_1,\ldots,f_M$ in a signal embedded in noise. We assume that the signal has the following structure:

$$\begin{aligned} x(t) &= y(t) + n(t) \\ &= \sum_{m=1}^{M} a_m(t)\exp(j(2\pi f_m t + \phi_m)) + n(t). \end{aligned} \tag{4}$$

Note that this model assumes that the signal is analytical; experimentally this is the case since data are obtained via passband (lock-in) sampling using the modulus and the phase of the acoustic wave. The time series is regularly sampled with time period T_s, so as to insure $1/T_s > 2f_{\max}$ where $f_{\max}$ stands for the bandwidth of the antialiasing filter used in the recording process.

The methods relies upon the identification of a signal subspace (respectively, noise subspace). At infinite signal-to-noise ratio (SNR), this subspace is spanned by the M spectral lines; its identification requires us to build an observation space of dimension $M+1$ at least. To this end, we introduce a set of Q K-dimensional observation vectors ($Q>K>M$):

$$\begin{aligned} &\mathbf{Y}(t_j) = [y(t_j), y(t_j+1), y(t_j+2), \ldots, y(t_j+(K-1))]^T \\ &\qquad j = 1,...,Q. \end{aligned} \tag{5}$$

In this expression the t_j are chosen such that the $\mathbf{Y}$ vectors are linearly independent and the sampling frequency is set to $T_s=1$, and the term *frequency* will refer to *normalized* frequency (i.e., the actual frequency, divided by $F_s=1/T_s$).

The signal $\mathbf{Y}$ is completely defined by its constitutive frequencies and corresponding amplitudes (deterministic but unknown)

$$\begin{aligned} &F = \{f_1,\ldots,f_M\} \\ &\mathbf{A} = [a_1\exp(j\phi_1),\ldots,a_M\exp(j\phi_M)]^T \\ &\mathbf{S}(\mathbf{F}) = [\mathbf{S}_1,\ldots,\mathbf{S}_M] \\ &\qquad = \begin{pmatrix} 1 & \exp(2\pi f_1) & \cdots & \exp(2\pi(K-1)f_1 \\ \vdots & & & \vdots \\ 1 & \exp(2\pi f_M) & \cdots & \exp(2\pi(K-1)f_M \end{pmatrix}^T, \\ &\mathbf{Y} = S(\mathbf{F})\mathbf{A}. \end{aligned} \tag{6}$$

Under the infinite SNR assumption, the signal subspace is completely identified and the M frequencies are easily calculated.[16] However, we wish to build a method that handles situations of low SNR, for which the available observations are $\mathbf{X}=\mathbf{Y}+\mathbf{N}$, where $\mathbf{N}$ is an additive noise vector, statistically independent of the signal. We assume that it is complex, white Gaussian circular and iid (independent increment identically distributed) with (unknown) variance σ^2.

The distribution function of a k-dimensional vector $\mathbf{N}$ defined by

$$\mathbf{N}(t)=[n(t),n(t+1),n(t+2),\ldots,n(t+(K-1))]^T, \quad (7)$$

reads

$$p(\mathbf{N})=\frac{1}{(\sqrt{2\pi}\sigma)^K}\exp\left(-\frac{|\mathbf{N}|^2}{2\sigma^2}\right). \quad (8)$$

In the following, the term *observation* refers to a set of Q K-dimensional vectors $\mathbf{X}$:

$$\mathbf{X}(t_j)=[x(t_j),x(t_j+1),x(t_j+2),\ldots,x(t_j+(K-1))]^T, \quad j=1,...,Q. \quad (9)$$

One observation is thus associated with a time window of length $K+Q-1$—the actual duration of one observation is $T_{\text{obs}}=(Q+K-2)T_s$. Although we aim at identifying frequency fluctuations, we assume that the frequency parameters $f_1,\ldots,f_M$ remain constant over the duration of one observation. This implies that the physical signal is oversampled. Under the assumption that the noise process is iid, the log likelihood of a subset of Q' independant realizations of $\mathbf{X}$ is given by

$$\mathcal{L}(\mathcal{P})=-KQ'\log(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{q=1}^{Q'}|\mathbf{X}(q)-\mathbf{S}(\mathbf{F})\mathbf{A}(q)|^2. \quad (10)$$

According to the maximum likelihood principle, the set $\mathcal{P}$ of parameters must be chosen in order to maximize expression (10).

2. Reduced expression

Minimizing (10) jointly for all the parameters is usually untractable. Most authors propose a separate maximization for each of the parameters. For our application, the spectral components (i.e., $\mathbf{A}$ and $\mathbf{F}$) are the relevant variables. We first maximize with respect to $\mathbf{A}$ and derive an expression for the optimal $\mathbf{F}$; σ^2 is estimated independently.

The value of vector $\mathbf{A}$ which minimizes the norm $|\mathbf{X}(q)-\mathbf{S}(\mathbf{F})\mathbf{A}(q)|^2$ is easily obtained

$$\mathbf{A}(q)=(\mathbf{S}^+\mathbf{S})^{-1}\mathbf{S}^+(\mathbf{F})\mathbf{X}(q). \quad (11)$$

Note that the ''signal only'' vector $\mathbf{Y}=\mathbf{X}-\mathbf{N}$ appears to be the orthogonal projection of $\mathbf{X}$ on the signal subspace spanned by the row vectors of $\mathbf{S}$

$$\mathbf{Y}=\mathbf{S}(\mathbf{F})\cdot\mathbf{A}(q)=\mathbf{S}((\mathbf{S}^+\mathbf{S})^{-1}\mathbf{S}^+(\mathbf{F})\mathbf{X}=\Pi_s(\mathbf{F})\mathbf{X}, \quad (12)$$

where $\Pi_s(\mathbf{F})$ stands for the parametric projector on the signal subspace.[21] Let $\Pi_n(\mathbf{F})=\mathbf{I}-\Pi_s(\mathbf{F})$ be the noise subspace, $\mathbf{I}$ is the identity matrix. By substituting (12) and using the definition of $\Pi_n(\mathbf{F})$ in the expression of the log likelihood (10), one gets the following simplified expression to minimize:

$$L(\mathbf{F})=\frac{1}{\sigma^2}\sum_{q=1}^{Q}|\Pi_n(\mathbf{F})\mathbf{X}(q)|^2. \quad (13)$$

Using the properties of the trace operator (hereafter denoted Tr) and those of the projection matrix $\Pi_n(\mathbf{F})$, the maximum likelihood estimation of $\mathbf{F}$ takes the common form: minimize

$$L(\mathbf{F})=\frac{Q}{\sigma^2}\text{Tr}[\Pi_n(\mathbf{F})\hat{\mathbf{R}}_x], \quad (14)$$

where $\hat{\mathbf{R}}_x$ is an estimate of the correlation matrix $\mathbf{R}_x$ of the vector process $\mathbf{X}(q)$. Minimizing $L(\mathbf{F})$ in (14) leads to the exact value $\mathbf{F}_{\text{ML}}$, which has the maximum likelihood. The actual vectors $\mathbf{X}$ are obtained by time shift over the recorded time series; therefore, the observed vectors may not be considered as being corrupted by independent realizations of the noise process, as some ''time integration'' is performed in the estimation of $\mathbf{R}_x$. The consequences and interest of such smoothing have been studied by Clergeot and Tressens,[17] and Ouamri,[20] in the frame of array processing (in this context, time integration becomes ''spatial smoothing''). In the remainder of this paper, the development is based on Eq. (14), no matter how $\mathbf{R}_x$ is estimated; see the Appendix for the practical implementation.

3. Approximate maximum likelihood

Equation (14) is still too complicated to be solved analytically in a simple way. A minimization can be easily performed if $L(\mathbf{F})$ has a quadratic dependance in $\mathbf{S}$.[17] Let $\mathbf{R}_y$ be the correlation matrix of the signal vectors $\mathbf{Y}(q)$, the assumption that signal and noise are independent allow to establish the following equalities:

$$\hat{\mathbf{R}}_x=\mathbf{R}_y+\hat{\sigma}^2\mathbf{I}, \quad (15)$$

$$\mathbf{R}_y=\mathbf{SPS}^+, \quad (16)$$

$$\mathbf{P}=\mathcal{E}[\mathbf{AA}^+], \quad (17)$$

where $\mathcal{E}$ stands for the mathematical expectation.

Substituting in Eq. (14) leads to

$$L(\mathbf{F})=\frac{Q}{\hat{\sigma}^2}\text{Tr}[\Pi_n(\mathbf{F})\mathbf{S}\hat{\mathbf{P}}\mathbf{S}^+]. \quad (18)$$

Clergeot and Tressens[17] propose a second-order approximation of $L(\mathbf{F})$

$$L_{\text{AML}}(\mathbf{F})=\frac{Q}{\hat{\sigma}^2}\text{Tr}[\hat{\Pi}_n\mathbf{S}(\mathbf{F})\hat{\mathbf{P}}\mathbf{S}^+(\mathbf{F})], \quad (19)$$

in which $\hat{\Pi}_n$ is estimated by computing the projector spanned by the $(K-M)$ smallest eigenvalues of the estimated covariance matrix $\hat{\mathbf{R}}_x$. They prove that this approach leads to more reliable estimates of $\mathbf{F}$ at low SNR, and that the minimization of L_{AML} is asymptotically efficient. In practice, the following set of equations is used:

$$\hat{\sigma}^2=\frac{1}{K-M}\text{Tr}(\hat{\Pi}_n\hat{\mathbf{R}}_x), \quad (20)$$

$$\Pi_s(\mathbf{F})=\mathbf{S}(\mathbf{F})(\mathbf{S}^+(\mathbf{F})\cdot\mathbf{S}(\mathbf{F}))^{-1}\cdot\mathbf{S}^+(\mathbf{F}), \quad (21)$$

$$\mathbf{S}(\mathbf{F})\cdot\hat{\mathbf{P}}\cdot\mathbf{S}^+(\mathbf{F})=\Pi_s(\mathbf{F})(\hat{\mathbf{R}}_x-\sigma^2\mathbf{I})\cdot\Pi_s(\mathbf{F}). \quad (22)$$

The approximately quadratic dependence of L_{AML} in $\mathbf{S}(\mathbf{F})$, allows a fast convergence of the minimization algo-

rithm by using a simple Newton–Gauss algorithm

$$\mathbf{F}(k+1)=\mathbf{F}(k)-\mathbf{H}^{-1}\cdot\mathbf{grad}(L_{\mathrm{AML}})\big|_{\mathbf{F}=\mathbf{F}(k)}, \tag{23}$$

where k stands for the iteration step in the minimization process, **grad** and **H** are the gradient and Hessian, respectively (see the expressions in the Appendix).

Let us emphasize that in the previous expression, the recursion applies at a given time, and converges towards the frequency estimate for a unique observation window. The manner in which consecutive estimates are obtained by sliding the window in time and integrating new measurements is addressed in the next paragraph.

4. Combining new measurements and estimates

In this section, it is assumed that new measurements do not allow by themselves the derivation of a good estimate. The variance of such an estimate varies as $1/T_{\mathrm{obs}}\cong(K+Q-1)^{-1}$, whereas integrating new measurements to this estimate allow a better estimation. Let $\hat{\mathbf{F}}(t)$ be an estimate of **F** at time t, and $\mathcal{N}(\hat{\mathbf{F}}(t),\,\Gamma(t))$ its probability density, assumed to be normal with variance $\Gamma(t)$.[22] If a linear evolution model is known for $\mathbf{F}(t)$, one has

$$\mathbf{F}(t+1)=\mathbf{M}\mathbf{F}(t)+\varepsilon(t), \tag{24}$$

$$p_{t+1|t}(\mathbf{F})=\mathcal{N}(\mathbf{M}\hat{\mathbf{F}}(t),\mathbf{M}\Gamma(t)\mathbf{M}^{+}+\mathbf{R}_{\varepsilon}), \tag{25}$$

where **M** is the evolution matrix, ε is a perturbation term, which is statistically independent from **F**, and $\mathbf{R}_{\varepsilon}$ is its covariance matrix. $p_{t+1|t}$ is the probability density function that can be derived for time $t+1$, if the observations are made until time t only. As such an evolution equation is usually unknown, **M** will be set to the identity matrix in the rest of the paper (see Michel[19]) for a detailed discussion). Applying the Bayes rule over conditional probabilities gives

$$p_{t+1|t+1}(\mathbf{F})=\frac{p_{t+1|t}(\mathbf{F})\cdot p_{t+1}(\mathbf{X}|\mathbf{F})}{p_{t+1}(\mathbf{X})}. \tag{26}$$

Noting that $\log(p_{t+1}(\mathbf{X}|\mathbf{F}))$ is the log-likelihood function for which a reduced expression has been derived in the previous section, one gets, after all reductions and identifications, the simple following expressions:

$$\hat{\mathbf{F}}(t+1|t)=\hat{\mathbf{F}}(t), \tag{27}$$

$$\Gamma(t+1|t)=\Gamma(t)+\mathbf{R}_{\varepsilon}, \tag{28}$$

$$\Gamma(t+1)^{-1}=\mathbf{H}+\Gamma(t+1|t)^{-1}, \tag{29}$$

$$\hat{\mathbf{F}}(t+1|t+1)=\hat{\mathbf{F}}(t+1)=\hat{\mathbf{F}}(t+1|t)-\Gamma(t+1)^{-1}\cdot\mathbf{grad}, \tag{30}$$

where it can be shown that the gradient function has the same expression as in the previous section. The expression for grad is given in the Appendix. $\mathbf{R}_{\varepsilon}$ is an unknown matrix which will be practically set to $v^2\mathbf{I}$, where v^2 will be tuned in order to allow the algorithm to take slight changes in **F** into account. Furthermore, it is interesting that the set of expression above expresses a generalized Kalman filter for estimating **F** (in the sense that it relies upon second-order expansion of the log-likelihood functions). The statistical convergence properties and numerical efficiency of these approaches are described in the work of Michel and Clergeot[18] and Michel.[19]

IV. EXPERIMENTAL RESULTS

We first describe the simple case of a particle settling in a fluid at rest. It is well adapted to the reassigned spectrogram method because the acoustic signal has a good SNR and a slow frequency modulation. We show that it allows the extraction of the subtle interaction between the falling particle and its wake. We then study the more complicated case of the motion of a particle embedded in a turbulent flow, where the dynamics of motion is much faster and the SNR is poor. We show that the AML method is well suited.

A. The settling sphere

1. Motivation and experimental setup

When a particle is released in a fluid at rest, its developing motion creates a wake. The particle velocity is then set by the balance between buoyancy forces and drag, and additional subtle effects: first, ''added mass'' corrections because the particles ''pushes'' the fluid, and second, a ''history'' force because the wake reacts back on the particle. Formally, one can write the equation of motion as[23–25]

$$\left(m_p+\frac{1}{2}m_f\right)\frac{\mathrm{d}\mathbf{v}_p}{\mathrm{d}t}=(m_p-m_f)\mathbf{g}-\tfrac{1}{2}\pi a^2\rho_f\|\mathbf{v}_p\|\mathbf{v}_p c_D(\mathrm{Re})+\mathbf{F}_{\mathrm{history}}, \tag{31}$$

where m_p is the particle mass, m_f is the mass of a fluid particle of the same size, v_p is the particle velocity, **g** is the acceleration of gravity, a is the sphere radius, ρ_f is the fluid density, c_D is the static empiric drag coefficient, Re is the Reynolds number $\mathrm{Re}=2av_p/\nu$ (ν is the fluid's kinematic viscosity), and $\mathbf{F}_{\mathrm{history}}$ is the so-called history force. In this expression, the drag coefficient is usually obtained from measurement of the forces acting on a body at rest in a hydrodynamic tunnel. The history term, however, is largely unknown. Analytic expression can only be derived in the limit of small Reynolds numbers (less than 10) and cannot be applied for real flow configurations (e.g., multiphase flows) where $\mathrm{Re}\gg 1$.

We perform measurements of the motion of a settling sphere, with the aim of evaluating the influence of the history forces. We use a water tank of size 1.1×0.75 m and depth 0.65 m, filled with water at rest (Fig. 2). The bead is held by a pair of tweezers, 5 cm below the transducers. It is released at time $t=0$ without initial velocity and its trajectory is about 50 cm long. The data acquisition is started before the bead is released in order to capture the onset of motion.

2. Results

Let us use as a first example the fall of a steel bead, 0.8 mm in diameter. A typical time series is shown in Fig. 3. The frequency of the scattered sound is low at the beginning of the fall and increases as the particle accelerates. The amplitude decreases as the distance to the transducer also increases. The Doppler shift during the bead motion is detected

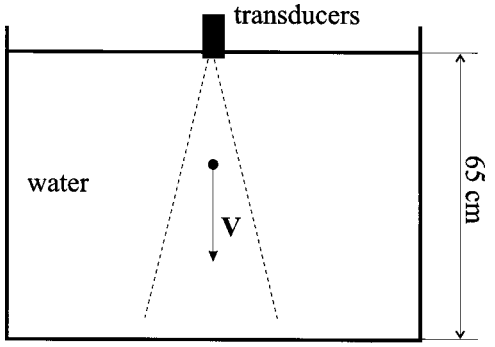


FIG. 2. Experimental setup in the case of the settling sphere. The bead is released 5 centimeters under the transducers, without initial velocity.

using the spectrogram representation and a subsequent reassignment scheme. The simple spectrogram and reassigned version are shown in Fig. 4. The reassignment technique drastically improves the localization of the energy in the time-frequency plane. In this case, the image processing step computes $v_p(t)$ as the line of maxima. The precision of the overall measurement depends on two factors: first on the intrinsic precision of the reassignment method and second on the dispersion of the measurements (the reproducibility of the bead motion over several experiments). The intrinsic precision of the reassignment method has been empirically studied using synthetic signals modeling the particle dynamics plus a noise that mimics the experimental data. We observed that for our choice of parameters (a time-frequency picture with 256×256 pixels) the rms precision is about one half-pixel both in time-and frequency directions. The method thus allows a precise analysis of the dynamics of the fall; we describe below two sets of experiment that illustrate the potential of the reassignment technique.

First, we show in Fig. 5 the velocity of a 1-mm steel bead (averaged over ten falls) together with two numerical simulations based on Eq. (31), first without the memory

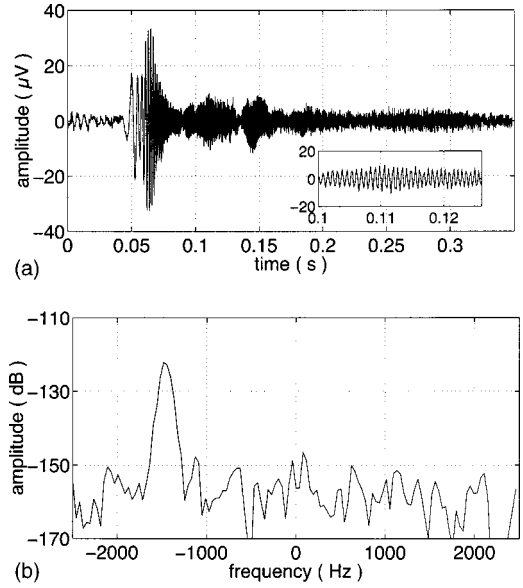


FIG. 3. Data from a steel bead (diameter 0.8 mm) settling in water at rest. (a) Typical time series; (b) power spectral density of the inset figure. On the *x* axis, zero corresponds to the emission frequency.

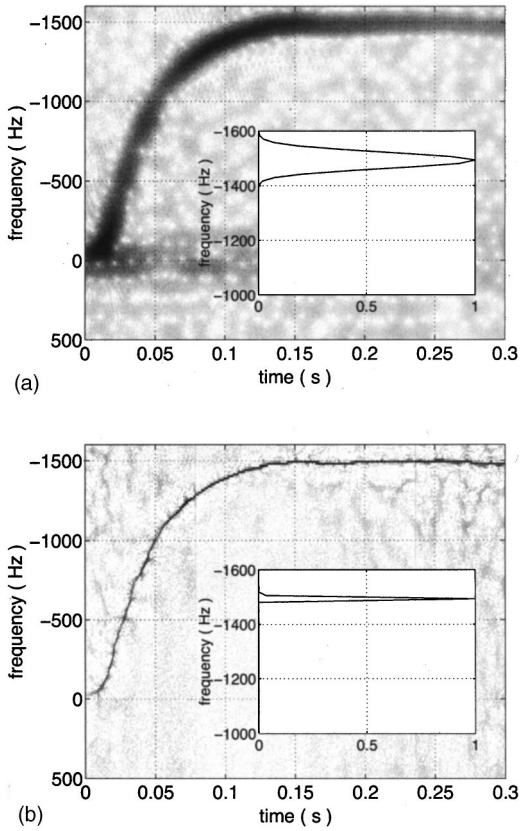


FIG. 4. (a) Spectrogram of the backscattered sound, after heterodyne detection. (b) Reassigned spectrogram. In each figure the inset shows a normalized cross section of the spectrogram. The algorithm is that of the tfrrsp function of the MATLAB time-frequency toolbox (Ref. 32). To get rid of the spectral components at zero frequency due to the coupling between transducers and at small frequencies around zero due to slow motion of the water surface, we use a high-pass fifth-order Butterworth filter of cutoff frequency 25 Hz (corresponding to a velocity of 5 mm/s). Data of an 0.8-mm steel bead settling in water at rest.

force and second with the expression of the memory force derived at low Reynolds numbers (called the Stokes memory term, as in Maxey and Riley[23]). The precision of the detection technique is sufficient for the measured profile to be compared to the simulated curves and to draw physical con-

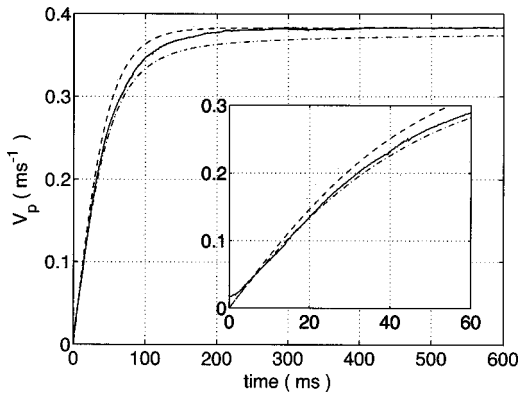


FIG. 5. Velocity measurement of a steel bead of diameter 1 mm (solid line), compared to numerical simulations without memory force (dashed) and with Stokes memory (dash-dotted). The inset shows an enlargement near the onset of motion. The Reynolds number, based on the limit velocity is 430. The sphere velocity profile results from averaging $n=10$ successive experiments.

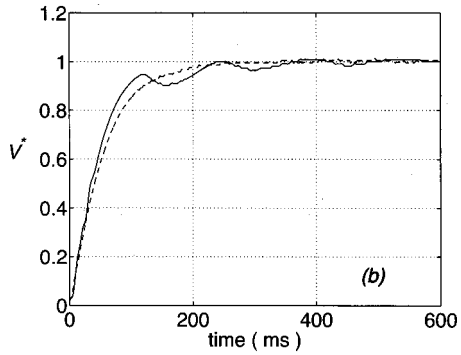


FIG. 6. Fall of a tungsten carbide sphere $D=1$ mm (dashed) compared to a glass bead $D=2$ mm (solid), at Re$\sim$400. The velocity is nondimensionalized by the limit velocity. Curves are not averaged over several experiments.

clusions about the hydrodynamical forces. At early times, the trajectory is close to the simulation with memory force. This is due to the diffusion away from the bead surface of the vorticity generated at the boundary.[23–25] However, as the instantaneous Reynolds number increases, the curve deviates from this simple regime: vorticity is advected into the wake. Memory is progressively lost and the sphere reaches a terminal velocity in a finite time, as does the simulation without memory.

The measurement and signal-processing techniques are then tested on a more nonstationary motion, as in the case of a bead whose density is closer to that of the fluid. In this situation a stronger interaction is expected between the particle motion and the development of its wake. Formally, this traces back to differences in the effective inertial mass and buoyancy mass of the particle—see Eq. (31). In Fig. 6, we show the velocity variation for a light glass sphere (density 2.48) compared to a tungsten bead (density 14.8). We observe that the velocity of the glass oscillates before reaching a constant terminal value, whereas the other particle has a regular acceleration. In the case of light beads the hydrodynamic forces may be large enough to overcome the gravity and change the sign of the acceleration. This is linked with the nonstationarity of the wake, as vortex shedding is known to occur for Reynolds number above critical (Re$_c\sim$250).

B. Turbulent flow: Lagrangian velocity measurement

1. Motivation and experimental setup

Although extensive Eulerian measurements (i.e., performed at a fixed point in space) are available in turbulent flows,[26] Lagrangian data are quite scarce.[6,27] Numerical simulations are the main source of Lagrangian information,[2] together with transposition to the time domain of Kolmogorov's mean-field theory, developed for spatial velocity increments. Of particular importance is the Lagrangian velocity autocorrelation function, which plays a central role in the modeling of turbulent dispersion (pollutants, contaminants, flames, cloud formation, etc.). The main models assume an exponential decay

$$\langle v(t)v(t+\tau)\rangle_t = u_{\rm rms}^2 e^{-\tau/T_L}, \tag{32}$$

with T_L a time scale characteristic of energy injection. This has been suggested by low Reynolds number studies[2] but not measured directly in a fully turbulent flow. Such a functional form is in agreement with arguments based on Kolmogorov phenomenology: a velocity increment over a time lag τ should depend only on the power injection per unit mas ϵ and the time lag τ itself. Dimensionally, the only possibility is $\langle (v(t+\tau)-v(t))^2\rangle_t \propto (\epsilon\tau)$, as in Eq. (32), when the time lag τ is much smaller than the time scale of energy injection T_L.

In order to obtain an adequate measurement of the autocorrelation function, one should be able to track tracer particles for times up to the largest time scales that characterize the forcing of the flow. To this end, it is helpful to choose a confined turbulent flow and to study it in the region where the mean velocity is almost null. The turbulent flow that we use is a von Kármán swirling flow: the water is set into motion by two coaxial counter-rotating disks in a cylindrical tank [Fig. 7(a)]. The flow pattern (averaged in time) is made of a differential toroidal velocity and a poloidal recirculation, as shown in Fig. 7(b). In our experiments, the cylindrical vessel has a linear dimension of the order of 20 cm, and with disks rotating at 5 Hz and above, the Reynolds number Re $=2\pi R^2 f/\nu$ exceeds 10^6—where $\nu=0.8910^{-6}\,\mathrm{m^2\,s^{-1}}$ is the water kinematic viscosity. In the core region considered here, the generated turbulence approximates isotropic and homogeneous conditions and its statistical characteristics are comparable to that of the largest wind tunnel experiments.[28–30] The time scales of the flow range from the period of rotation of the driving disks, about 100 ms, down to the Kolmogorov dissipative time scale, equal about 0.5 ms.

For the acoustic measurement, we use the same transducers as in the previous experiments, at emitting frequency 2.5 MHz. They are located at the walls, in the median plane, in order to span the central region of the flow—Fig. 7(b) shows the geometry of the measurement zone. The cylinder and the surface of the disks are covered by 3-cm Ureol 5073A and 6414B by CIBA Corp. Its relative density is 1.1 and the sound velocity is 1460 $\mathrm{m\,s^{-1}}$ so that its acoustic

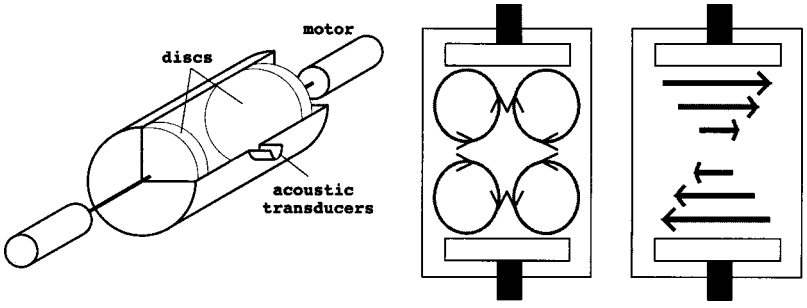


FIG. 7. Experimental setup. The inner radius of the cylinder is 10 cm (disk radius $R=9.5$ cm) and the distance between the disks is 18 cm. They are driven by two 1-kW motors at a constant rotation frequency equal to $f=18$ Hz. The acoustic transducers are placed 18 cm off-axis, in order to increase the volume of the measurement region. The structure of the time-averaged flow is shown in the right two pictures. Each disk creates an azimuthal motion (center picture) and generates a poloidal recirculation (Ref. 33).

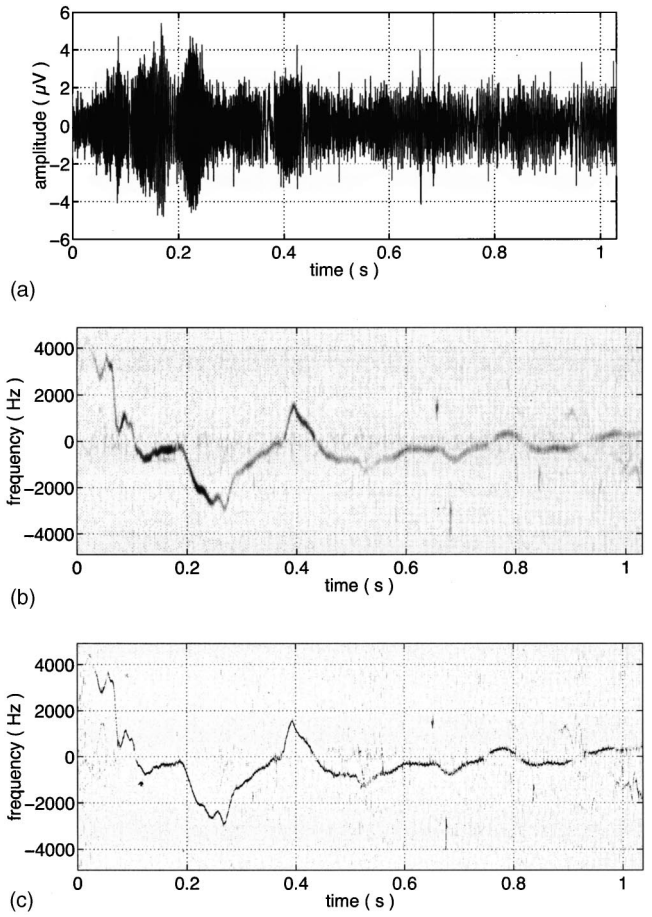


FIG. 8. Sound scattered by a 2-mm diameter PP bead in a turbulent flow at $\mathrm{Re}=10^6$. (a) Typical time series; (b) and (c) corresponding spectrogram and reassigned spectrogram.

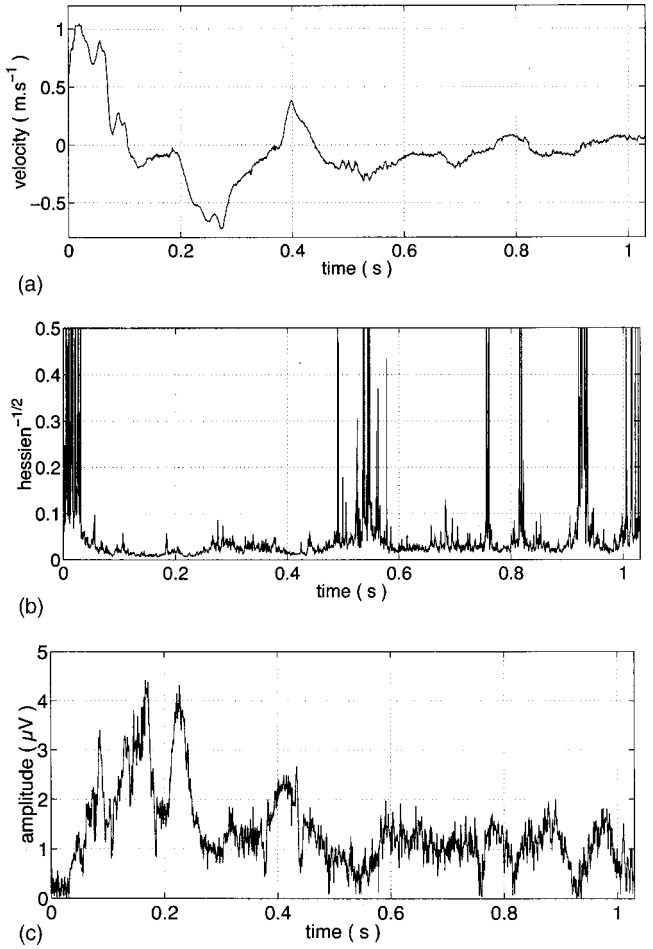


FIG. 9. Output of the AML algorithm for the motion of a 2-mm-diameter PP bead in a turbulent flow at $\mathrm{Re}=10^6$. (a) Velocity; (b) corresponding inverse square root of the Hessian; (c) amplitude of the source (the rms value of the noise is 0.9 μV). AML algorithm parameters: $M=1$, $K=7$, $Q=13$, $v^2=10^{-5}$.

impedance is close to that of the water, drastically reducing the reflections at the interface water/ureol compared to water/steel. The attenuation at 2.5 MHz is about 6 dB per cm, so that with a 3-cm layer, the total reflection is reduced by a factor of 60. The particle is a polypropylene (PP) sphere of radius 1 mm and relative density 0.9.

If the turbulent motion of the particle is to be tracked, one needs to measure the Doppler shift of the sound scattered with a time resolution of the order of $\delta t \sim 1$ ms. This means that the Doppler shift must be of the order of 1 kilohertz. If one imposes a velocity precision better than 5%, the frequency resolution must be of the order of $\delta f \sim 50$ Hz, so that $\delta t\, \delta f < 1$. The uncertainty principle implies that time-frequency techniques without *a priori* information about the scattered signal cannot be used in this problem. We show in the next section that a tracking scheme using our high-resolution parametric detection is well adapted.

2. Results

Figure 8 shows (a) a time series; (b) the corresponding spectrogram; and (c) the reassigned spectrogram for the case of one particle in the ultrasonic beam. The signal to noise ratio is very poor, typically less than 6 dB [to give an idea, in Fig. 8(a) the bead enters the ultrasonic beam at $t \sim 20$ ms]. The time-frequency pictures show the trajectory of the particle but the low SNR prevents it from being easily extracted. In particular, the trajectory in the reassigned picture becomes quite lacunar and extracting it would require sophisticated (and CPU-greedy) image processing techniques.

The result of the AML algorithm is plotted in Fig. 9. The extracted frequency modulation is of course within the estimation in the spectrogram as in Fig. 8(b), but one observes that fine variations in the velocity of the bead are now detected. The algorithm also provides an estimate of the amplitude of the source [Fig. 9(c)]. It can be seen that there is a strong amplitude modulation and that the SNR is at most 6 dB and may become less than 0 dB. As the Hessian is related to the Fisher information matrix,[31] its inverse square root is linked with the variance of the estimation: a large value of the Hessian indicates an accurate estimation of the modulation frequency and, hence, of the bead velocity. The inverse square root of the Hessian is plotted in Fig. 9(b): very large values are calculated in the absence of a bead in the measurement volume at the beginning and end of the time series (as a signature of the mismatch between the model which is composed of at least one source and the reality: no source). Local lower values (typically less than 0.1) are observed when the variance on the estimation is small. Spurious effects are generated when the frequency modulation ap-

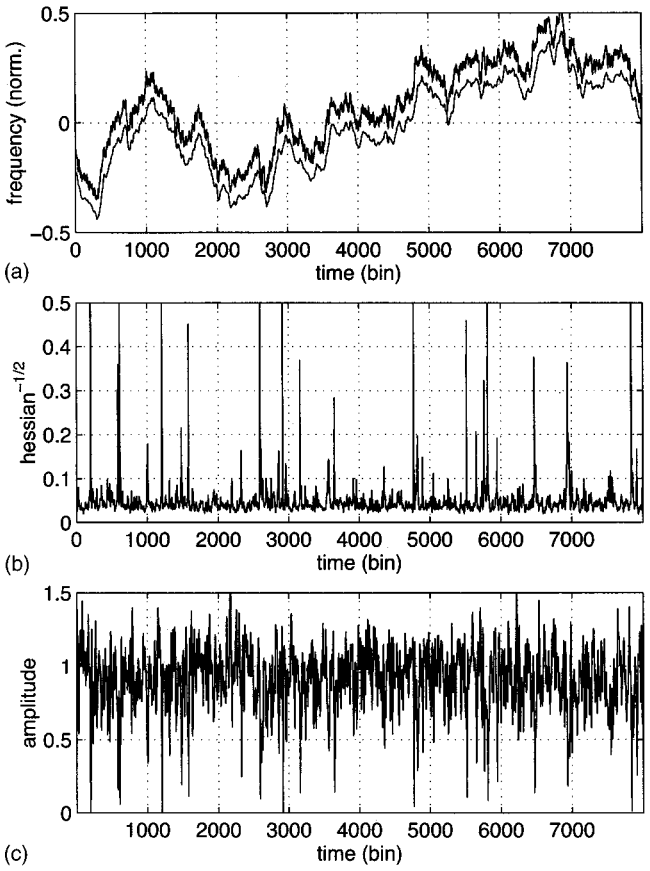


FIG. 10. Test of the AML algorithm on a synthetic signal having the same frequency modulation and SNR as the turbulence data. (a): original (upper curve) and detected (lower curve) frequency modulation—the vertical shift is added for clarity. (b) corresponding inverse square root of the Hessian; (c) amplitude of the source (original set to 1). AML algorithm parameters: $M=1$, $K=7$, $Q=13$, $v^2=10^{-5}$.

proaches zero as the Hessian also becomes very small because of the filtering operation done in order to get rid of the coupling part of the signal. Finally, one observes that the Hessian decreases as the signal-to-noise ratio increases (see at time 0.55 s).

In order to test the algorithm, we have applied it to a synthetic signal x_s

$$x_s(j)=\exp\left(2i\pi\sum_{p=1}^{j}\phi(p)\right)+n(j). \tag{33}$$

The frequency modulation $\phi(p)$ is chosen to mimic the turbulence data: it is made of a Gaussian random variable with a $1/f^2$ time spectrum. The complex white noise n is a random variable with a Gaussian amplitude and a uniformly distributed phase; its variance is equal to 1 so that the SNR in this test is 0 dB. The signal is processed by the AML algorithm with its adjustable parameters tuned to the same values as in the turbulence detection ($M=1$, $K=7$, $Q=13$, $v^2=10^{-5}$). The output is shown in Fig. 10. One observes that the detection is very good. The effective low-pass filtering is due to the finite size of the observation window in the estimation of the frequency modulation and to the generalized Kalman filter used in the tracking of its time fluctuations. The inverse Hessian varies as in the turbulence data, with a value mostly lower than 0.1, indicating that the frequency modulation is

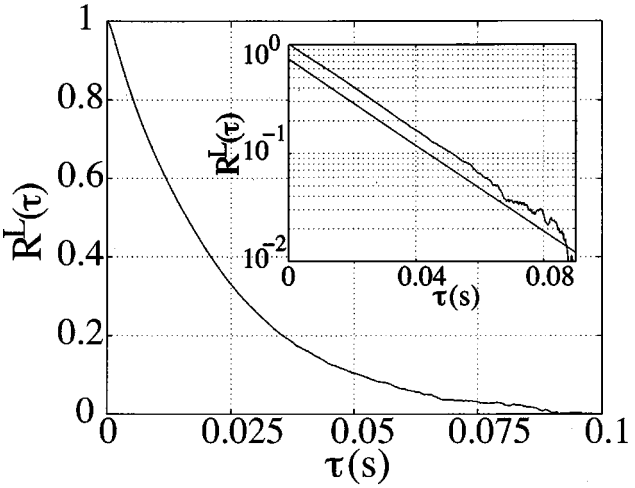


FIG. 11. Velocity autocorrelation coefficient, for a rotation rate of 430 rpm. A best exponential fit is $\rho_v^L(\tau)=1.03e^{-45.7\tau}$. It is shown, slightly shifted for clarity, as the linear curve in the inset.

well estimated. This test establishes the performance of the AML algorithm and validates the velocity measurement presented above.

As a first validation of the entire particle-tracking technique as a Lagrangian measurement in turbulence, we compute the autocorrelation coefficient of the velocity fluctuations (one component only of the particle velocity is measured here)

$$\rho_v^L(\tau)=\frac{\langle v_x(t)v_x(t+\tau)\rangle_t}{v_{\mathrm{rms}}^2}. \tag{34}$$

As shown in Fig. 11, the decrease of the velocity correlation is remarkably well fitted by an exponential form. We obtain $\rho_v^L(\tau)=1.03e^{\tau/T_L}$ with $T_L=17$ ms. This behavior is in agreement with previous observations in direct numerical simulations of homogeneous isotropic turbulence [POPE]; the integral time T_L is characteristic of the flow forcing (energy injection).

V. CONCLUDING REMARKS

As can be seen in the previous section, both methods, time-frequency analysis and parametric spectral analysis, are suited for extracting the time-varying frequency modulation due to a Doppler effect. The domain of application of each method depends on the degree of nonstationarity and on the SNR.

For high SNR and weakly nonstationary signals, the time-frequency approach yields very good results. One drawback is the need of a second processing stage to extract the trajectory from the time-frequency picture. This stage may become increasingly difficult if there is more than one spectral component or if the SNR degrades. In both cases the quadratic nature of the algorithm produces interference patterns in the image: spurious clusters and a lacunar trajectory result. Another, more fundamental, limitation is that the length of the time window must be long enough to preserve an acceptable frequency resolution, even with the reassigned spectrogram. This limits the methods to weakly nonstationary signals.

For signals with a rapid frequency modulation, the AML spectral estimation is well suited, as long as the noise is near iid. The size of the time window can be decreased because of the parametric nature of the method, since *a priori* knowledge has been taken into account. The performance is further increased by the use of a Kalman-type filter. The drawback is the necessity to find a good dynamical model for the evolution of the spectral components. We have chosen here the simplest model which works well for our experiments, but the approach can be refined by increasing the number of parameters in order to consider more precisely the variation of the frequency. The AML algorithm also provides a quantitative estimation of the quality of the demodulation and the instantaneous power of the spectral component. It has the advantage to provide directly the frequency modulation as a function of time, in one stage.

The association of a Doppler acoustic technique with time-frequency or high-resolution parametric signal-processing algorithms yields a resolved measurement of single-particle velocities. We have validated this approach in two extreme cases; one example where an object moves in a quiet fluid and one example where the object is set into motion by turbulence. In both cases, the motion of the particle is successfully tracked. This new technique has potential applications in the tracking of solid bodies entrained (or propelled) in complex flows at low or high Reynolds numbers.

ACKNOWLEDGMENTS

We are indebted to Pascal Metz for the development of the signal conditioning electronics. We thank Marc Moulin for his help in the design of the von Kármán setup, VERMON for continuous assistance in the development of the transducer array. This work is partially supported by ACI Grant No. 2226.

APPENDIX: AML ALGORITHM

(1) First step: calculate $\hat{\mathbf{R}}_x$ using the following expression:[19]

$$\hat{\mathbf{R}}_x=\frac{1}{2Q}\sum_{i=t+1}^{t+Q}(\mathbf{X}(i)\mathbf{X}(i)^T+\tilde{\mathbf{X}}(i)\tilde{\mathbf{X}}(i)^T), \tag{A1}$$

with

$$\tilde{\mathbf{X}}(i)=[x(i+K-1),x(i+K-2),\ldots,x(i)]^{*T}, \tag{A2}$$

where * stands for complex conjugate. $\tilde{\mathbf{X}}\mathbf{X}$.

(2) Second step: diagonalize $\hat{\mathbf{R}}_x$; one obtains the eigenvectors $(\mathbf{V}_i)_{i=1,\ldots,K}$ and eigenvalues $(\lambda_i)_{i=1,\ldots,K}$ sorted in decreasing order.

(3) Third step: Compute $\hat{\Pi}_n$ and $\hat{\sigma}^2$, using the set of equations

$$\hat{\Pi}_n=\sum_{i=M+1}^{K}\mathbf{V}_i\mathbf{V}_i^T, \tag{A3}$$

$$\hat{\sigma}^2=\frac{1}{K-M}\mathrm{Tr}(\hat{\Pi}_n\hat{\mathbf{R}}_x)=\frac{1}{K-M}\sum_{i=M+1}^{K}\lambda_i. \tag{A4}$$

(4) Fourth step: choose $\mathbf{F}=\hat{\mathbf{F}}(t)$ as a candidate value. Compute **grad** and **H** using[18]

$$\mathbf{grad}=\frac{2Q}{\sigma^2}\mathrm{R_e}\{\mathrm{Diag}(\mathbf{S}'^{+}(\mathbf{F})\cdot\Pi_n(\mathbf{F})\cdot\hat{\Pi}_n\cdot\mathbf{S}(\mathbf{F})\cdot\hat{\mathbf{P}})\}, \tag{A5}$$

$$\mathbf{H}=\frac{2Q}{\sigma^2}\mathrm{R_e}\{\mathrm{Diag}((\mathbf{S}'^{+}(\mathbf{F})\cdot\Pi_n(\mathbf{F})\cdot\hat{\Pi}_n\cdot\Pi_n(\mathbf{F})\cdot\mathbf{S}'))\star\hat{\mathbf{P}}^*\}, \tag{A6}$$

where the operator $\star$ stands for the term to term matrix multiplication, and $\mathbf{P}^*$ is the conjugate of **P**, and

$$\mathbf{S}'=\left[\frac{d\mathbf{S}_1}{df_1},\ldots,\frac{d\mathbf{S}_M}{df_M}\right]^T. \tag{A7}$$

(5) Fifth step: using Eqs. (27)–(30), compute $\hat{\mathbf{F}}(t+1)$ and $\Gamma(t+1)$.

The initialization of the algorithm is done by either (i) setting an initial value of $\mathbf{F}(1)$ or (ii) estimating this value using the maxima of the amplitude of the FFT of a small window of signal (of length 64 or 128 samples) and using the iterative algorithm described in Sec. III B 3 to converge towards $\mathbf{F}(1)$.

For example, the extracted velocity of Fig. 9 is obtained by starting at the maximum energy of the signal and applying the algorithm forward and backward in time. The algorithm is stopped, as the mean of the inverse square root of the Hessian over a window of size 400 samples exceeds 0.5 for more than 400 samples.

[1] G. O. Fountain, D. V. Khakhar, I. Mezic, and J. M. Ottino, J. Fluid Mech. **417**, 265–301 (2000).
[2] P. K. Yeung and S. B. Pope, J. Fluid Mech. **207**, 531–586 (1989).
[3] K. D. Squires and J. K. Eaton, Phys. Fluids A **3**, 130–143 (1991).
[4] P. K. Yeung, Phys. Fluids **6**, 3416–3428 (1994).
[5] P. K. Yeung, Phys. Fluids **9**, 2981–2990 (1997).
[6] M. Virant and T. Dracos, Meas. Sci. Technol. **8**, 1539–1552 (1997).
[7] G. A. Voth, K. Satyanarayan, and E. Bodenschatz, Phys. Fluids **10**, 2268–2279 (1998).
[8] O. F. Bay and I. Güler, J. Med. Syst. **23**, 77–84 (1999).
[9] P. Flandrin, *Time-Frequency/Time Scale Analysis*, Wavelet Analysis and its Application, Vol. 10 (Academic, New York, 1998).
[10] L. Cohen, *Time-Frequency Analysis*, edited by A. V. Oppenheim (Prentice-Hall, Englewood Cliffs, NJ, 1995).
[11] G. Gaunard and H. Uberall, J. Acoust. Soc. Am. **73**, 1–12 (1983).
[12] P. D. Thorne, T. J. Brudner, and K. R. Waters, J. Acoust. Soc. Am. **95**, 2478–2487 (1994).
[13] B. T. Hefner and P. L. Marston, J. Acoust. Soc. Am. **107**, 1930–1936 (2000).
[14] R. Gendrin and C. de Villedary, Ann. Telecommun. **35**, 122–130 (1979).
[15] F. Auger and P. Flandrin, IEEE Trans. Signal Process. **43**, 1068–1089 (1995).
[16] S. Kay, *Modern Spectral Estimation, Theory and Application*, Prentice-Hall Signal Processing Series, edited by A. V. Oppenheim (Prentice-Hall, Englewood Cliffs, NJ, 1988).
[17] H. Clergeot and S. Tressens, in ICASSP'90, Albuquerque, NM (1990).
[18] O. Michel and H. Clergeot, in ICASSP'91, Toronto, Canada (1991), pp. 1277–1280.
[19] O. Michel, Ph.D. thesis, Université Paris XI, Orsay (1991).
[20] A. Ouamri, Ph.D. thesis, Université Paris-Sud, Orsay (1986).
[21] It is straightforward to establish that $\Pi_s\mathbf{Y}=\mathbf{Y}$ for any vector **Y** lying in the signal subspace. Here, "parametric projector" must be understood as the projector calculated for the vector of frequencies **F**.

[22] This is generally not the case, but this assumption remains valid as long as the log likelihood is well approximated by its second-order expansion around $\hat{\mathbf{F}}$.
[23] M. R. Maxey and J. J. Riley, Phys. Fluids **26**, 883–889 (1983).
[24] C. J. Lawrence and R. Mei, J. Fluid Mech. **283**, 307–327 (1995).
[25] N. Mordant and J.-F. Pinton, Eur. Phys. J. B **18**, 343–352 (2000).
[26] U. Frisch, *Turbulence, the Legacy of A. N. Kolmogorov* (Cambridge University Press, Cambridge, 1995).
[27] A. L. Porta, G. A. Voth, A. M. Crawford, J. Alexander, and E. Bodenschatz, Nature (London) **409**, 1017–1019 (2001).
[28] J.-F. Pinton and R. Labb, J. Phys. II **4**, 1461–1468 (1994).
[29] J. Maurer, P. Tabeling, and G. Zocchi, Europhys. Lett. **26**, 31–36 (1994).
[30] A. Arneodo, C. Baudet, F. Belin, R. Benzi, B. Castaing, B. Chabaud, R. Chavarria, S. Ciliberto, R. Camussi, F. Chill *et al.*, Europhys. Lett. **34**, 411–416 (1996).
[31] L. Scharf, *Statistical Signal Processing; Detection, Estimation and Time Series Analysis* (Addison-Wesley, Reading, MA, 1991).
[32] Trademark, The Mathworks Company. The time-frequency toolbox can be downloaded at http://crttsn.univ-nantes.fr/~auger/tftbftp.html
[33] L. Marié, J. Burguete, A. Chiffaudel, F. Daviaud, D. Ericher, C. Gasqquet, F. Petielis, S. Fauve, M. Bourgoin, M. Moulin, P. Odier, J.-F. Pinton, A. Guigon, JB Luciani, F. Namer, J. Léorat, in ''Dynamo and dynamics'', Proceedings of the NATO Advanced Research Workshop, Cargèse (France), August 2000, NATO Science Series II Vol. 26; P. Chossat, D. Armbruster and I. Opera Eds. Kluwer Acad. Pub., Dordrech, The Netherlands (2001).

The performance of matched-field track-before-detect methods using shallow-water Pacific data

Stacy L. Tantum,[a] Loren W. Nolte,[b] Jeffrey L. Krolik,[c] and Kerem Harmanci[d]
Department of Electrical Engineering, Box 90291, Duke University, Durham, North Carolina 27708-0291

(Received 27 August 1998; revised 20 December 2001; accepted 29 April 2002)

Matched-field track-before-detect processing, which extends the concept of matched-field processing to include modeling of the source dynamics, has recently emerged as a promising approach for maintaining the track of a moving source. In this paper, optimal Bayesian and minimum variance beamforming track-before-detect algorithms which incorporate *a priori* knowledge of the source dynamics in addition to the underlying uncertainties in the ocean environment are presented. A Markov model is utilized for the source motion as a means of capturing the stochastic nature of the source dynamics without assuming uniform motion. In addition, the relationship between optimal Bayesian track-before-detect processing and minimum variance track-before-detect beamforming is examined, revealing how an optimal tracking philosophy may be used to guide the modification of existing beamforming techniques to incorporate track-before-detect capabilities. Further, the benefits of implementing an optimal approach over conventional methods are illustrated through application of these methods to shallow-water Pacific data collected as part of the SWellEX-1 experiment. The results show that incorporating Markovian dynamics for the source motion provides marked improvement in the ability to maintain target track without the use of a uniform velocity hypothesis. [DOI: 10.1121/1.1489435]

PACS numbers: 43.30.Wi, 43.60.Gk [SAC-B]

I. INTRODUCTION

The passive detection, localization, and tracking of submerged acoustic sources are areas of research which have received considerable attention due to their importance in civilian and military surveillance applications. Performing these tasks in shallow coastal waters tends to be challenging because in addition to the complicated multipath nature of the acoustic propagation, there often are many uncertainties concerning the parameters which describe the acoustic environment. Matched-field processing exploits the complicated multipath environment and provides an effective method for source localization when the propagation model is completely known.[1] However, matched-field processing requires accurate information about the source parameters and propagation characteristics of the water channel in order to accurately model the propagation and predict the acoustic field. If the assumptions regarding the sound speed profile or other environmental parameters are inaccurate, the predicted fields will be incorrect and performance is degraded. This is the classic sensitivity of matched-field processing to environmental mismatch which has been explored in several papers.[2,3] Thus, a significant effort has been made to evaluate the effects of environmental uncertainty on matched-field processing techniques and to develop algorithms which are robust to this uncertainty.[4] The optimum uncertain field processor[5] (OUFP) and the minimum variance beamformer with environmental perturbation constraints[6] are among the matched-field processing algorithms which exhibit robustness to environmental uncertainty.

Source motion is also a factor which degrades the performance of matched-field processing algorithms if it is not taken into consideration.[7] Matched-field track-before-detect processing has recently emerged as a promising approach for handling a moving source. This technique extends the concept of matched-field processing to include modeling of the source dynamics. However, many of the algorithms initially developed assume a uniformly moving target.[8,9] This source motion model does not allow the source to change speed or direction over the integration period of the track, which imposes limits on its tracking abilities. For example, if a source is moving at a constant speed but not in a constant direction over the track, the uniform motion model is not applicable and the performance of these algorithms will be poor. An alternative matched-field tracking approach which does not assume uniform source motion is the optimum uncertain field tracking algorithm[10] (OUFTA). The OUFTA introduces a Markov model for the source motion as a means of capturing the stochastic nature of the source dynamics without assuming uniform motion. The Markov model assumes *a priori* that from one observation to the next, the source moves only in the neighborhood of its previous location.

In this paper, an optimal minimum variance track-before-detect (MV-TBD) method is presented which exploits a recently derived equivalence between adaptive minimum variance beamforming and maximum likelihood source localization. Although track-before-detect processing has been previously used with matched-field beamformers, the resulting methods have not yielded an optimal approach for incor-

[a]Electronic mail: slt@ee.duke.edu
[b]Electronic mail: lwn@ee.duke.edu
[c]Electronic mail: jk@ee.duke.edu
[d]Electronic mail: harmanci@ee.duke.edu

TABLE I. Symbol definitions for Bayesian techniques.

Symbol	Definition
$\boldsymbol{\Psi}$	Acoustic environment parameters
$\mathbf{S}$	Source position
$\boldsymbol{\Phi}$	Source parameters (amplitude and phase)
$\mathbf{r}(t)$	Received time domain signal
$\mathbf{s}(t)$	Received time domain source signal
$\mathbf{n}(t)$	Additive time domain noise
$\mathbf{P}(\mathbf{r})$	Received frequency domain signal
A	Frequency domain signal complex amplitude
σ_A^2	Source amplitude variance
$\mathbf{H}(\mathbf{S},\boldsymbol{\Psi})$	Replica field
$\mathbf{N}$	Frequency domain diffuse noise
σ_N^2	Noise variance
$\mathbf{Q}$	Noise covariance matrix

porating source dynamics. Further, the resulting MV-TBD method is related to the OUFTA, thus illustrating how an optimal tracking philosophy can be used to guide the modification of existing beamforming approaches.

A second contribution of this paper is the application of the OUFTA and MV-TBD processors to real data. Data collected during the SWellEX-1 experiment[11] has been processed using the OUFTA and MV-TBD beamformers, and the results are presented here along with comparisons to the results achieved with conventional processing methods. This paper shows that implementing optimal tracking processors provides marked improvement in source tracking capabilities over conventional processing methods.

The organization of the paper is as follows. The signal and noise models are first presented in Sec. II, and the Bayesian tracking algorithms are then explained in Sec. III. Track-before-detect (TBD) beamforming and its relationship to the optimal tracking framework is presented in Sec. IV. Section V contains the results of applying TBD beamforming and the tracking algorithms to the SWellEX-1 data, and Sec. VI concludes with a discussion of the results.

II. SIGNAL AND NOISE MODELS

A. Notation

This section presents the models used for optimal Bayesian and minimum variance beamforming track-before-detect algorithms. Historically, the matched-field processing and array processing literature has followed different notational conventions. In order to remain consistent with previously published work, the descriptions of the Bayesian and minimum variance models will retain their conventional notations. The symbols used by each of these approaches and their definitions are listed in Tables I and II. In addition, Table III lists the symbols for which there is a correspondence between the two techniques and provides comments clarifying their relationships.

B. Modeling for Bayesian *a posteriori* probability methods

The optimum uncertain field processor[5] (OUFP) and the optimum uncertain field tracking algorithm[10] (OUFTA) are Bayesian *a posteriori* probability methods used for source localization and tracking. These methods calculate the *a posteriori* probability of the parameter(s) to be estimated given the observed signal, $\mathbf{r}_i(t)$, where the subscript i denotes a time index assigned to sequential observations. The signal emitted by the source is assumed to be a narrowband sinusoid with a known frequency, f_0 [Hz], and the observed time domain signal across the receiving array is assumed to consist of the received source signal, $\mathbf{s}_i(\mathbf{S}_i,\boldsymbol{\Psi},\boldsymbol{\Phi}_i,t)$, plus additive noise, $\mathbf{n}(t)$;

$$\mathbf{r}_i(t)=\mathbf{s}_i(\mathbf{S}_i,\boldsymbol{\Psi},\boldsymbol{\Phi}_i,t)+\mathbf{n}(t). \tag{1}$$

The received source signal at each array element is a function of the source position relative to the array element, $\mathbf{S}_i$, the propagation parameters associated with ocean acoustic waveguide, $\boldsymbol{\Psi}$, and the amplitude and phase parameters of the acoustic source, $\boldsymbol{\Phi}_i$.

The frequency transform of the received signal is of the form

$$\mathbf{P}_i(\mathbf{r}_i)=A_i\mathbf{H}(\mathbf{S}_i,\boldsymbol{\Psi})+\mathbf{N}, \tag{2}$$

where A_i is a complex Gaussian random variable with parameter σ_A^2 associated with the source parameters $\boldsymbol{\Phi}_i$, and $\mathbf{H}(\mathbf{S}_i,\boldsymbol{\Psi})$ is the acoustic transfer function, or replica field in matched-field terminology, for the narrowband source located at the position $\mathbf{S}_i$ in the ocean $\boldsymbol{\Psi}$. Each of the observations $\mathbf{P}_i(\mathbf{r}_i)$ is assumed to contain additive zero-mean Gaussian noise $\mathbf{N}$ with a known spatial covariance matrix $\mathbf{Q}$.

TABLE II. Symbol definitions for minimum variance beamforming techniques.

Symbol	Definition
θ	Source position
s^o	Frequency domain signal complex amplitude
$\mathbf{x}$	Received data snapshot
$\mathbf{X}$	Frame of multiple snapshot data
$\mathbf{d}_\theta$	Source wave front vector
$\mathbf{n}$	Diffuse noise vector
σ_N^2	Diffuse noise variance
$\mathbf{R_n}$	Diffuse noise covariance matrix
$\mathbf{i}$	Interference wave front vector
$\boldsymbol{\Psi}$	Interference covariance matrix defined in the orthogonal complement subspace of $\mathbf{d}_\theta$
$\mathbf{e}$	Diffuse noise+interference
$\mathbf{C_e}$	Diffuse noise+interference covariance matrix
$\mathbf{T}_\theta$	Matrix whose columns define the orthogonal complement of $\mathbf{d}_\theta$

TABLE III. Symbol equivalences between Bayesian and minimum variance beamforming techniques.

OUFTA symbol	MV-TBD symbol	Comments
$\mathbf{S}$	θ	Exact equivalence
A	s^o	Random vs nonrandom
$\mathbf{H}(\mathbf{S},\boldsymbol{\Psi})$	$\mathbf{d}_\theta$	Exact equivalence
$\mathbf{P}(\mathbf{r})$	$\mathbf{x}$	Exact equivalance
$\mathbf{N}$	$\mathbf{e}$	Diffuse noise vs diffuse noise+interference
σ_N^2	σ_N^2	Exact equivalence
$\mathbf{Q}$	$\mathbf{C_e}$	Diffuse noise vs diffuse noise+interference

In this work, the noise is assumed to be isotropic, consequently $\mathbf{Q}=\sigma_N^2\mathbf{I}$.

When the source amplitude is assumed to be a complex Gaussian random variable with parameter σ_A^2, the probability density function of the observation given the source position and the environmental parameters is[5]

$$p(\mathbf{r}_i|\mathbf{S}_i,\boldsymbol{\Psi})=\frac{1}{E(\mathbf{S}_i,\boldsymbol{\Psi})+1}\exp\left(\frac{1/2|R(\mathbf{r}_i,\mathbf{S}_i,\boldsymbol{\Psi})|^2}{E(\mathbf{S}_i,\boldsymbol{\Psi})+1}\right), \tag{3}$$

where $E(\mathbf{S}_i,\boldsymbol{\Psi})$ is related to the energy in the replica field and $R(\mathbf{r}_i,\mathbf{S}_i,\boldsymbol{\Psi})$ is related to the correlation between the replica field and the observed field. The quantities $E(\mathbf{S}_i,\boldsymbol{\Psi})$ and $R(\mathbf{r}_i,\mathbf{S}_i,\boldsymbol{\Psi})$ are defined by

$$\begin{aligned}E(\mathbf{S}_i,\boldsymbol{\Psi})&=\sigma_A^2\mathbf{H}^\dagger(\mathbf{S}_i,\boldsymbol{\Psi})\mathbf{Q}^{-1}\mathbf{H}(\mathbf{S}_i,\boldsymbol{\Psi})\\&=\frac{\sigma_A^2}{\sigma_N^2}\mathbf{H}^\dagger(\mathbf{S}_i,\boldsymbol{\Psi})\mathbf{H}(\mathbf{S}_i,\boldsymbol{\Psi})\end{aligned} \tag{4}$$

and

$$\begin{aligned}R(\mathbf{r}_i,\mathbf{S}_i,\boldsymbol{\Psi})&=\sigma_A^2\mathbf{H}^\dagger(\mathbf{S}_i,\boldsymbol{\Psi})\mathbf{Q}^{-1}\mathbf{P}(\mathbf{r}_i)\\&=\frac{\sigma_A^2}{\sigma_N^2}\mathbf{H}^\dagger(\mathbf{S}_i,\boldsymbol{\Psi})\mathbf{P}(\mathbf{r}_i).\end{aligned} \tag{5}$$

In addition, the signal-to-noise ratio (SNR) is defined at the receivers and is given by

$$\begin{aligned}\mathrm{SNR}=E(\overline{\mathbf{S}},\overline{\boldsymbol{\Psi}})&=\sigma_A^2\mathbf{H}^\dagger(\overline{\mathbf{S}},\overline{\boldsymbol{\Psi}})\mathbf{Q}^{-1}\mathbf{H}(\overline{\mathbf{S}},\overline{\boldsymbol{\Psi}})\\&=\frac{\sigma_A^2}{\sigma_N^2}\mathbf{H}^\dagger(\overline{\mathbf{S}},\overline{\boldsymbol{\Psi}})\mathbf{H}(\overline{\mathbf{S}},\overline{\boldsymbol{\Psi}}).\end{aligned} \tag{6}$$

The OUFTA extends the OUFP to include modeling of the source motion in addition to uncertainties in the parameters affecting the received source signal. The source motion is modeled as a first-order finite-state discrete-time Markov process. This model fits well in the matched-field processing framework since the range and depth are already discretized in order to compute the ambiguity surfaces. With this source motion model, the distribution of the current source position is dependent only on the previous source position, and the joint probability of any source track reduces to the product of the marginal probabilities of the current source position given the previous position,

$$p(\mathcal{S}_k)=\prod_{i=1}^{k}p(\mathbf{S}_i|\mathbf{S}_{i-1}), \tag{7}$$

where $p(\mathbf{S}_1|\mathbf{S}_0)$ is the probability of the source starting in the initial position. Under this model, the nature of the source dynamics is captured by a transition matrix which contains the probabilities of the source moving from its location at the current observation to a neighboring location at the next observation. In effect, the transition probabilities describing the *a priori* knowledge of the source motion, $p(\mathbf{S}_i|\mathbf{S}_{i-1})$, permit movement of the source only within the neighborhood of its previous position. In addition, the random complex amplitude of the acoustic source is assumed to be independent from observation to observation.

C. A structured covariance matrix model

Most of the models on which source localization and tracking have been applied assume extensive knowledge of the structure of the noise and/or interference covariance matrix. This section describes a distinct data model, proposed in Ref. 12, in which the level of required prior knowledge on the noise and/or interference covariance structure is reduced. This model can be used to relate a parametric maximum likelihood (ML) source localization approach to minimum variance (MV) beamforming.

Consider M snapshots of the narrowband complex envelope of received data, measured by an array of N sensors, $\{\mathbf{x}_1,\ldots,\mathbf{x}_M\}$, described as

$$\mathbf{x}_m=s_m^o\mathbf{d}_\theta+\mathbf{i}_m+\mathbf{n}_m,\quad m=1,\ldots,M, \tag{8}$$

where $\mathbf{d}_\theta$ is the signal wave front given a single source located at θ, normalized so that $\mathbf{d}_\theta^\dagger\mathbf{d}_\theta=1$. The temporal samples of the signal, s_m^o, are treated as deterministic unknowns. The white noise, $\mathbf{n}_m$, and the interference, $\mathbf{i}_m$, are zero-mean, Gaussian vectors, independent of each other and temporally uncorrelated, where the interference is spatially correlated with unknown covariance. The white noise is spatially uncorrelated with covariance: $\mathbf{R_n}=\sigma_N^2\mathbf{I}$. The problem addressed here is to estimate the source location, θ. For all snapshots m, interference $\mathbf{i}_m$ and noise $\mathbf{n}_m$ are independent and identically distributed.

First, note that the unknown source location, signal temporal samples, and interference covariance cannot be jointly estimated without assuming a structure for the covariance matrix.[13] However, let $[\mathbf{d}_\theta\mathbf{T}_\theta]$ be a matrix whose columns form a complete orthonormal basis and decompose the interference, $\mathbf{i}_m$, as follows:

$$\mathbf{i}_m=\alpha_{m1}\mathbf{d}_\theta+\sum_{k=2}^{N}\alpha_{mk}\mathbf{t}_k, \tag{9}$$

where $\mathbf{T}_\theta\triangleq[\mathbf{t}_2,\ldots,\mathbf{t}_N]$ and $\alpha_{mk}\triangleq\mathbf{t}_k^\dagger\mathbf{i}_m$. Using Eq. (9), Eq. (8) can be rewritten as

$$\mathbf{x}_m=(s_m^o+\alpha_{m1})\mathbf{d}_\theta+\sum_{k=2}^{N}\alpha_{mk}\mathbf{t}_k+\mathbf{n}_m,\quad m=1,\ldots,M. \tag{10}$$

The coefficient of $\mathbf{d}_\theta$ consists of two terms, s_m^o and α_{m1}. The scalar s_m^o cannot be jointly estimated with the interference covariance, because it is a deterministic unknown parameter corrupted with a random variable, α_{m1}, whose variance is unknown. Therefore, this paper proposes a model where the terms s_m^o and α_{m1} are combined as $s_m\triangleq(s_m^o+\alpha_{m1})$. The new variable s_m is assumed to be deterministic unknown while the coefficients of the component of the interference in the orthogonal complement to the signal wave front, $(\alpha_{m2},\ldots,\alpha_{mN})$, are assumed to be Gaussian with unknown covariance, $\boldsymbol{\Psi}$. Thus, the model is reformulated as

$$\mathbf{x}_m=s_m\mathbf{d}_\theta+\mathbf{e}_m,\quad m=1,\ldots,M, \tag{11}$$

where $\{s_m\}_{m=1}^M$ are unknown parameters and $\{\mathbf{e}_m\}_{m=1}^M$ are independent zero-mean Gaussian vectors with covariance

$$\mathbf{C_e} = \mathbf{T}_\theta \boldsymbol{\Psi} \mathbf{T}_\theta^\dagger + \sigma_N^2 \mathbf{I}. \tag{12}$$

Note that, unless $\mathbf{i}_m^\dagger \mathbf{d}_\theta = 0$, s_m no longer represents the signal wave form. However, since the goal here is to estimate the signal direction, the s_m are simply nuisance parameters and as can be shown, this parametrization leads to a well-defined ML estimator. Thus to recap, the objective is to estimate the source location, θ, in the presence of the nuisance parameters, $\mathbf{s} \triangleq [s_1, \ldots, s_M]^T$ and $\boldsymbol{\Psi}$ from the measurements $X \triangleq [\mathbf{x}_1, \ldots, \mathbf{x}_M]$. In order to formulate a source motion model, the narrowband sensor data are grouped into temporal frames, each containing M snapshots. The nth frame is represented by

$$X_n = [\mathbf{x}_{1n}, \ldots, \mathbf{x}_{Mn}], \tag{13}$$

where $\mathbf{x}_{mn}$ is the signal at the mth snapshot of the nth frame. The signal and noise for each vector $\mathbf{x}_{mn}$ is described by Eq. (11). It is assumed that the source location vector within each frame is fixed but it changes from frame to frame. The description of a frame with fixed source location has been introduced to validate the wide sense stationarity property, necessary for the use of structured covariance matrix model. Thus, for each frame, a nuisance parameter that is necessary to estimate is the data covariance matrix. The source position at the nth frame, θ_n, is modeled as a first-order Markov process, i.e.,

$$p(\theta_1, \ldots, \theta_n) = \prod_{l=1}^{n} \Pr(\theta_l | \theta_{l-1}), \tag{14}$$

where $p(\theta_1, \ldots, \theta_n)$ is the joint probability density function of the source locations at frames $(\theta_1, \ldots, \theta_n)$, and the transition probability, $\Pr(\theta_1 | \theta_0)$, is assumed to be uniformly distributed in a known interval. The transition probabilities $\Pr(\theta_l | \theta_{l-1})$ for $l > 1$ are defined such that the movement of the source at the lth frame is permitted only within the neighborhood of the source position in the $(l-1)^{st}$ frame.

III. BAYESIAN TRACKING METHODS

Two Bayesian techniques are implemented to estimate the path taken by a moving source. The conventional approach is to concatenate the results of a series of independent source localizations. This approach, termed conventional processing, is quite fast; however it does not incorporate the *a priori* knowledge of the source dynamics or carry forth environmental information from observation to observation. The optimum uncertain field tracking algorithm[10] (OUFTA) implements an optimal approach to tracking a moving source in the presence of environmental uncertainty. The OUFTA determines the most likely source path after considering all the observed data. It also utilizes the nature of the source motion as well as the assumption of a constant environment when tracking the source. This enables the OUFTA to increase its knowledge of the uncertain parameters as more data is obtained, and improves its performance. These processing methods are described in further detail in Secs. III A and III B.

A. Conventional processing

The conventional approach to this problem is to perform a series of independent source localizations, ignoring the *a priori* knowledge of the source dynamics and the constant state of the unknown ocean environment during the observation time. For each observation, the ambiguity surface is computed by the chosen localization processor, and the estimated source position for that observation, $\hat{\mathbf{S}}_i$, is determined by choosing the value of $\mathbf{S}$ which maximizes the ambiguity surface. The estimate of the source trajectory for k observations, $\hat{\mathcal{S}}_k$, is then formed by concatenating the individual source location estimates,

$$\hat{\mathcal{S}}_k = \hat{\mathbf{S}}_1, \hat{\mathbf{S}}_2, \ldots, \hat{\mathbf{S}}_k. \tag{15}$$

For this work, the source localizations are performed using the OUFP.

The OUFP calculates $p(\mathbf{S}|\mathbf{r})$, the *a posteriori* probability density function of the source location $\mathbf{S}$ given the received signal $\mathbf{r}$. The OUFP is robust to environmental variability because it does not assume a particular propagation environment. Instead, it assumes a range of probable propagation environments and then integrates over the environmental uncertainty to calculate the *a posteriori* probability density function,

$$p(\mathbf{S}|\mathbf{r}) = \frac{\int_{\boldsymbol{\Psi}} p(\mathbf{r}|\mathbf{S}, \boldsymbol{\Psi}) p(\mathbf{S}|\boldsymbol{\Psi}) p(\boldsymbol{\Psi}) d\boldsymbol{\Psi}}{p(\mathbf{r})}. \tag{16}$$

Given the assumptions regarding the signal model, the conditional probability density function, or ambiguity surface, $p(\mathbf{S}|\mathbf{r})$, can be expressed as

$$p(\mathbf{S}|\mathbf{r}) = C(\mathbf{r}) p(\mathbf{S}) \int_{\boldsymbol{\Psi}} \frac{1}{E(\mathbf{S}, \boldsymbol{\Psi}) + 1} \times \exp\left(\frac{1/2 \, |R(\mathbf{r}, \mathbf{S}, \boldsymbol{\Psi})|^2}{E(\mathbf{S}, \boldsymbol{\Psi}) + 1} \right) p(\boldsymbol{\Psi}|\mathbf{S}) d\boldsymbol{\Psi}, \tag{17}$$

where $C(\mathbf{r})$ is a normalization constant chosen to make $p(\mathbf{S}|\mathbf{r})$ a proper probability density function; $\int_{\mathbf{S}} p(\mathbf{S}|\mathbf{r}) d\mathbf{S} = 1$. The *a priori* knowledge of the source position, $p(\mathbf{S})$, is assumed to be uniform over all possible source locations.

When $\boldsymbol{\Psi}$ contains a small number of environmental parameters, a brute force numerical integration is performed to evaluate the integral. However, when $\boldsymbol{\Psi}$ contains many uncertain parameters brute force integration is computationally prohibitive. In this case, the integral can be evaluated efficiently using Monte Carlo integration.[14]

B. The optimum uncertain field tracking algorithm

The optimum uncertain field tracking algorithm[10] (OUFTA) implements an optimal approach to the problem of tracking a moving source in the presence of environmental uncertainty and determines the most likely source path after considering all the observed data. By extending the OUFP to include modeling of the source motion in addition to the uncertainty in the ocean, the *a posteriori* conditional probability of the source path given the sequence of k data observations, $p(\mathcal{S}_k|\mathcal{R}_k)$, can be computed, and from this the MAP

estimate of the source track can be determined. It utilizes the nature of the source motion as well as the assumption that the environment does not change while the source is being tracked. This enables the OUFTA to increase its knowledge of the uncertain parameters as more data are obtained, and improves its performance.

Given the model for the source motion and the assumed independence of the source amplitude between observations, the joint conditional probability of the observed data given the source track and environment reduces to the product of the individual probabilities of the observation given the source position and the environment,

$$p(\mathcal{R}_k|\mathcal{S}_k,\boldsymbol{\Psi})=\prod_{i=1}^{k}\left[\int_{\Phi}p(\mathbf{r}_i|\mathbf{S}_i,\boldsymbol{\Psi},\Phi_i)p(\Phi)d\Phi\right]. \quad (18)$$

Given the assumed signal model, it is known that each of these individual probabilities can be evaluated analytically [Eq. (3)]. According to Bayes' rule, the desired *a posteriori* conditional probability can be expressed as

$$p(\mathcal{S}_k|\mathcal{R}_k)=\frac{\Pi_{i=1}^{k}p(\mathbf{S}_i|\mathbf{S}_{i-1})}{p(\mathcal{R}_k)}\int_{\boldsymbol{\Psi}}\left\{\prod_{i=1}^{k}p(\mathbf{r}_i|\mathbf{S}_i,\boldsymbol{\Psi})\right\}p(\boldsymbol{\Psi})d\boldsymbol{\Psi}. \quad (19)$$

The source path with the maximum *a posteriori* probability is then chosen as the estimate of the source track,

$$\hat{\mathcal{S}}_k=\max_{\mathcal{S}_k}p(\mathcal{S}_k|\mathcal{R}_k). \quad (20)$$

Since all the previous observations are taken into consideration when making the current estimate of the source track, this algorithm has the ability to correct previous errors and refine the track estimate.

Ideally, it is desirable to compute the *a posteriori* probability of every possible source track; however the computational intensity of this task increases exponentially with the number of observations. In principle, this makes implementation of the OUFTA computationally prohibitive. The concepts behind the Viterbi algorithm[15,16] can be applied to reduce the computational complexity to linear growth with the number of observations and make implementation of this algorithm feasible. Using the Viterbi algorithm, the number of surviving source tracks remaining after each observation is equal to the number of possible source positions. This is accomplished by retaining only the one most probable path leading to each possible state, or source position, after every observation. The MAP estimate of the source path given the sequence of observed data can be found by solving the equivalent maximization problem of

$$\mathcal{S}_k=\max_{\mathcal{S}_k}\left\{\sum_{i=1}^{k}\ln\,p(\mathbf{S}_i|\mathbf{S}_{i-1})+\ln\int_{\boldsymbol{\Psi}}\left[\prod_{i=1}^{k}p(\mathbf{r}_i|\mathbf{S}_i,\boldsymbol{\Psi})\right]p(\boldsymbol{\Psi}|\mathcal{S}_k)d\boldsymbol{\Psi}\right\}. \quad (21)$$

This is the maximization problem addressed by the Viterbi algorithm. This expression can be interpreted as a weight which is assigned to each surviving path, where the first term is dependent upon the movement of the source between observations and the second term is dependent upon the position of the source at each observation.

IV. TRACK-BEFORE-DETECT BEAMFORMING

An important application of existing beamforming methods is to optimally incorporate *a priori* knowledge of the source motion, extending current beamformers to include source tracking capabilities. The minimum variance track-before-detect (MV-TBD) beamformer is developed in Sec. IV A, and the relationship between this beamformer and the OUFTA is then explained in Sec. IV B.

A. The minimum variance track-before-detect beamformer

The minimum variance (MV) beamformer has been widely applied in matched field processing as a means of suppressing ambiguous sidelobes by minimizing the output noise power while maintaining the integrity of the original signal.[17] The objective here is to optimally incorporate *a priori* knowledge of the source dynamics into the MV beamformer, extending the effective integration time of the processor.[18]

The likelihood function for the current source position θ_m using the model of Eq. (11)–(14) can be calculated by

$$L_{\mathrm{TBD}}(\theta_m)=p(\theta_m|X_m,\ldots,X_{m-L+1},\hat{\Phi}_m,\ldots,\hat{\Phi}_{m-L+1}), \quad (22)$$

where X_i represents an M-snapshot frame of data and $\hat{\Phi}_i$ represents the maximum likelihood estimate (MLE) of the noise and source parameters. Given the source track, the data frames are assumed to be independent with probability densities $p(X_k|\theta_k,\hat{\Phi}_k)$. In addition, it is assumed that the source has been present for the previous L data frames and its motion can be described by a Markov model. Consequently, the source position is a discrete random variable with Markovian dynamics such that

$$p(X_m,\ldots,X_{m-L+1},\theta_m,\ldots,\theta_{m-L+1})=\prod_{k=m-L+1}^{m}p(X_k|\theta_k,\hat{\Phi}_k)p(\theta_k|\theta_{k-1}), \quad (23)$$

where the initial probabilities $p(\theta_{m-L+1}|\theta_{m-L})$ are defined to be uniformly distributed over all possible source locations. The transition probabilities describing the source dynamics, $p(\theta_k|\theta_{k-1})$, permit movement of the source only within the neighborhood of its previous position.

Calculation of the likelihood function $L_{\mathrm{TBD}}(\theta_m)$ requires summing over all possible tracks. This is a computationally intensive task, however it can be computed recursively using forward variables,

$$\alpha_k(\theta^{(i)})=\left\{\sum_{b=1}^{B}\alpha_{k-1}(\theta^{(b)})p(\theta^{(i)}|\theta^{(b)})\right\}p(X_k|\theta_k^{(i)},\hat{\Phi}_k), \quad (24)$$

where i and b are indices representing the source location parameters of range and depth, B is the total number of possible source locations, and $k=m-L+2,\ldots,m$ is a time index. The forward variables are proportional to the probability

of the data frame given the source position and the MLE of the noise and source parameters,

$$\alpha_{m-L+1}(\theta^{(i)}) \propto p(X_{m-L+1}|\theta^{(i)}_{m-L+1}, \hat{\Phi}_{,-L+1}). \tag{25}$$

Harmanci *et al.* have shown that when a structured noise field is assumed the probability of the observation given the source position and MLE of the noise and source parameters in a known ocean is proportional to the output of the MV beamformer raised to the power of the number of snapshots,[12]

$$p(X_k|\theta^{(i)}_k, \hat{\Phi}_k) \propto \hat{Z}_{mv}(\theta^{(i)})^M. \tag{26}$$

Using this relationship, the log of the forward variables can be calculated by

$$\log\ \alpha_k(\theta^{(i)}) \propto \log\left\{ \sum_{b=1}^{B} \alpha_{k-1}(\theta^{(b)}) p(\theta^{(i)}|\theta^{(b)}) \right\} + M\ \log\ \hat{Z}_{mv}(\theta^{(i)}_k; X_k). \tag{27}$$

Therefore, the track-before-detect MV processor is

$$L_{\mathrm{TBD}}(\theta_m) = \log\ \alpha_m(\theta). \tag{28}$$

The wideband version is attained by averaging the logs of the single frequency MV beamformer outputs. This can be accomplished by replacing $M\ \log\ \hat{Z}_{mv}(\theta^{(i)}_k; X_k)$ with $M\Sigma_{k=1}^{K} \log\ \hat{Z}_{mv}(\theta^{(i)}_k; \omega_k)$.

B. Relating the OUFTA and the MV-TBD beamformer

The OUFTA and the MV-TBD beamformer were originally derived under different assumptions, and in order to show the relationship between the two algorithms these differences must be addressed. The most obvious difference between these algorithms is the quantity which is estimated; the OUFTA estimates the entire source track, $\mathcal{S}_k = \mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_k$, while the MV-TBD beamformer estimates the current source position, $\mathbf{S}_k$. Presently, the OUFTA calculates the joint probability of the source track given all the received data. Taking the marginal of this joint probability by integrating over all the previous source positions will provide the probability of the current source location given all the received data. The MAP estimate of the current source position can then be determined from these probabilities. The resulting expression is

$$p(\mathbf{S}_k|\mathcal{R}_k) = \int_{\mathbf{S}_1} \cdots \int_{\mathbf{S}_{k-1}} \frac{\Pi_{i=1}^{k} p(\mathbf{S}_i|\mathbf{S}_{i-1})}{p(\mathcal{R}_k)} \times \int_{\mathbf{\Psi}} \left\{ \prod_{i=1}^{k} p(\mathbf{r}_i|\mathbf{S}_i, \mathbf{\Psi}) \right\} p(\mathbf{\Psi}) d\mathbf{\Psi} d\mathbf{S}_{k-1} \ldots d\mathbf{S}_1, \tag{29}$$

where

$$p(\mathbf{r}_i|\mathbf{S}_i, \mathbf{\Psi}) = \int_{\Phi} p(\mathbf{r}_i|\mathbf{S}_i, \mathbf{\Psi}, \Phi_i) p(\Phi_i) d\Phi. \tag{30}$$

Some additional assumptions are also necessary in order to show the relationship between the OUFTA and the MV-TBD beamformer. The first assumption is that the environmental parameters represented by $\mathbf{\Psi}$ are independent across observations. This changes the expression from an integral over a product to a product of integrals, and eliminates the dependence on $\mathbf{\Psi}$. The second assumption is that the MLE of the source parameters, $\hat{\Phi}$, is a sufficiently accurate estimate and can be substituted for the actual parameters without degrading the performance due to mismatch, eliminating the integration over the uncertain source parameters. Applying these assumptions, the OUFTA becomes

$$p(\mathbf{S}_k|\mathcal{R}_k) = \int_{\mathbf{S}_1} \cdots \int_{\mathbf{S}_{k-1}} \frac{\Pi_{i=1}^{k} p(\mathbf{S}_i|\mathbf{S}_{i-1})}{p(\mathcal{R}_k)} \times \left\{ \prod_{i=1}^{k} p(\mathbf{r}_i|\mathbf{S}_i, \hat{\Phi}) \right\} d\mathbf{S}_{k-1} \ldots d\mathbf{S}_1, \tag{31}$$

where

$$p(\mathbf{r}_i|\mathbf{S}_i, \hat{\Phi}) = \int_{\mathbf{\Psi}} p(\mathbf{r}_i|\mathbf{S}_i, \mathbf{\Psi}, \hat{\Phi}) p(\mathbf{\Psi}) d\mathbf{\Psi}. \tag{32}$$

It is important to note that the OUFTA assumes knowledge of the noise covariance while the MV-TBD beamformer estimates the noise covariance given a sufficiently large number of data snapshots.

The final assumption is that the processor has a finite memory, limiting the integration to only the previous L source positions. Under this assumption, the final modified OUFTA formulation is

$$p(\mathbf{S}_k|\mathbf{r}_{k-L+1}, \ldots, \mathbf{r}_k) = \frac{\int_{\mathbf{S}_{k-L+1}} \cdots \int_{\mathbf{S}_{k-1}} \Pi_{i=k-L+1}^{k} p(\mathbf{r}_i|\mathbf{S}_i, \hat{\Phi}) p(\mathbf{S}_i|\mathbf{S}_{i-1}) d\mathbf{S}_{k-1} \ldots d\mathbf{S}_{k-L+1}}{p(\mathbf{r}_{k-L+1}, \ldots, \mathbf{r}_k)}. \tag{33}$$

The MV-TBD makes this assumption as a practical consideration. When an infinite integration time is implemented, the strongest source dominates the ambiguity surface, masking any weaker sources which may be present. A finite integration time enables the processor to track weaker sources in addition to the strongest source. However, the OUFTA was not designed to track multiple sources, and the optimal approach to this problem may not be simply limiting the integration time.

Finally, recall the relationship between the probability of

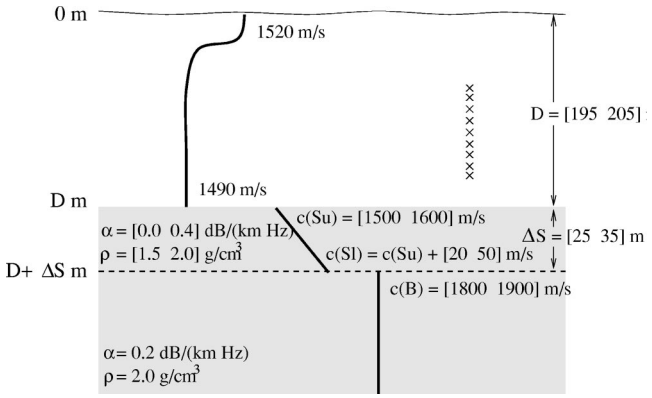


FIG. 1. Environmental configuration for the SWellEX-1 experiment.

the observation given the source position and the MLE of the noise and source parameters in a known ocean and the MV beamformer output [Eq. (26)],

$$p(X_k|\theta_k^{(i)},\hat{\Phi}_k)\propto\hat{Z}_{mv}(\theta^{(i)})^M. \tag{34}$$

From this, it can be seen that the output of the MV beamformer is proportional to the kernel of the modified OUFTA. Therefore, the tracking framework provided by the OUFTA can be used to guide the modification of existing beamformers to incorporate source tracking capabilities.

V. EXPERIMENTAL RESULTS

The SWellEX-1 experiment was conducted off the San Diego coast near Point Loma in August 1993. This site is a shallow water channel approximately 200 m in depth with a relatively flat bottom and a downward refracting sound speed profile. Figure 1 depicts the ocean environment in this region as well as the bounds placed on the seven uncertain environmental parameters. In addition, the perturbations on the sound speed in the water column are modeled by four empirical orthogonal functions, for a total of eleven uncertain parameters. The bounds placed on the uncertain parameters are taken from physically realistic values which have been experimentally determined.[11]

The data were collected at a 48-element vertical line array (VLA) with interelement spacing of 1.875 m suspended about 10 m above the ocean floor. Throughout the course of this experiment, the towed acoustic source moved away from the VLA at about 1.5 m/s and maintained a nearly constant depth around 80 m. The observations are formed by taking a 8192-point fast Fourier transform of the time sampled data ($f_s=1.5$ kHz) with 50% overlap. This procedure yields approximately 22 points of frequency domain data per minute.

The replica fields are computed by using normal mode theory to solve the acoustic wave equation. This method expresses the pressure field in terms of a normal mode expansion, and then solves for the eigenfunctions and eigenvalues which are solutions to the wave equation and satisfy the boundary conditions.[19] The total acoustic pressure field is then calculated as the weighted sum of the contributions from each mode, where the weighting of the *m*th mode is proportional to the amplitude of that mode at the source

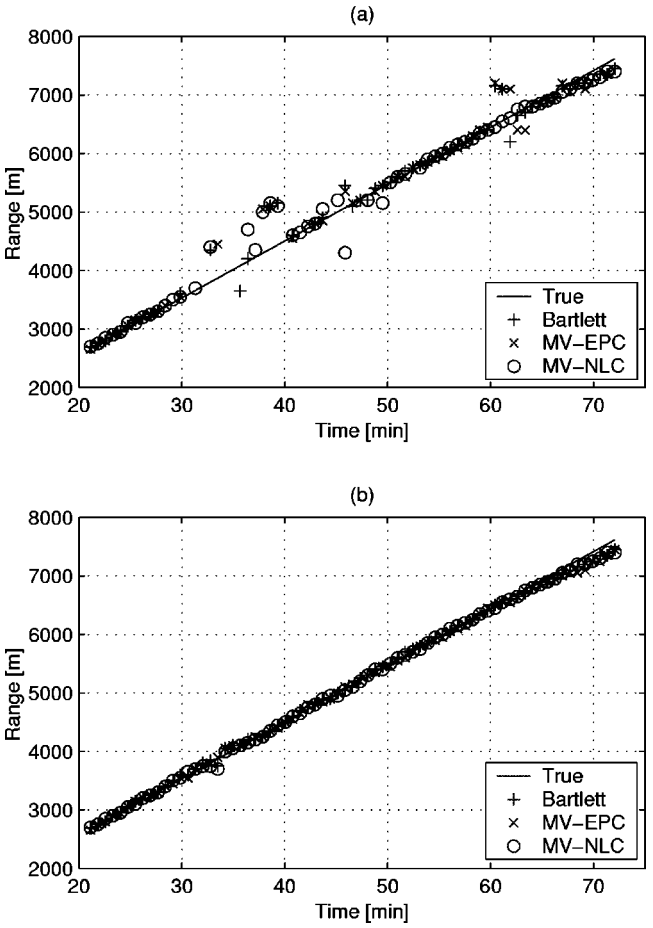


FIG. 2. SWellEX-1 experiment wideband range tracking performance for: (a) conventional beamforming, (b) track-before-detect beamforming with a 20 frame memory.

depth. For an isotropic point source with a known frequency f [Hz], the result is a Sturm–Liouville eigenvalue problem with an infinite number of solutions, each characterized by an eigenfunction and eigenvalue. The eigenvalues, or acoustic wave numbers, k_m, are all distinct and the eigenfunctions, or acoustic modes, $\Phi_m(z)$, form an orthonormal set. The normal mode program KRAKEN efficiently and accurately calculates the eigenfunctions, or acoustic modes, and eigenvalues, or acoustic wave numbers, for an ocean acoustic waveguide.[20,21] Once the acoustic wave numbers, k_m, and modes, $\Phi_m(z)$, are determined by KRAKEN, the complex-valued acoustic pressure field generated by a source at range r_s and depth z_s in a range independent ocean is calculated from

$$p(z;r_s,z_s)=p_0\sum_{m=1}^{M}\Phi_m(z_s)\Phi_m(z)\frac{e^{ik_m r_s}}{\sqrt{k_m r_s}}, \tag{35}$$

where z is the depth of the receiver, p_0 is the pressure generated by the source at $r_s=1$ m, and M is the number of propagating acoustic modes.

The results of processing a segment (J229-GI track) of the data collected as part of the SWellEX-1 experiment are presented here. The results obtained when these data are processed using wideband TBD beamforming are presented in Sec. V A. Conventional processing and the OUFTA are then

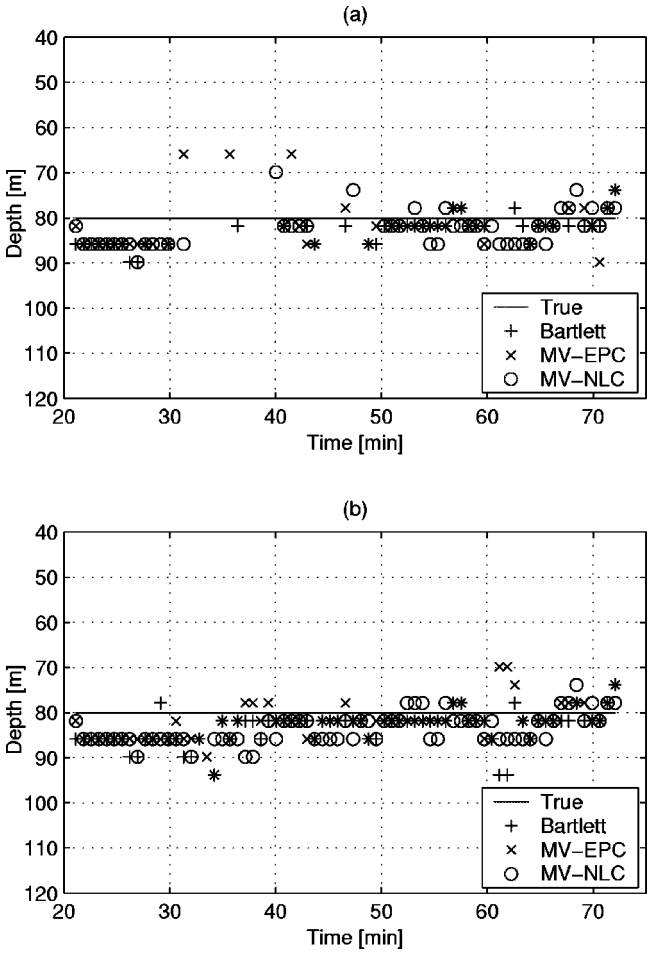


FIG. 3. SWellEX-1 experiment wideband depth tracking performance for: (a) conventional beamforming, (b) track-before-detect beamforming with a 20 frame memory.

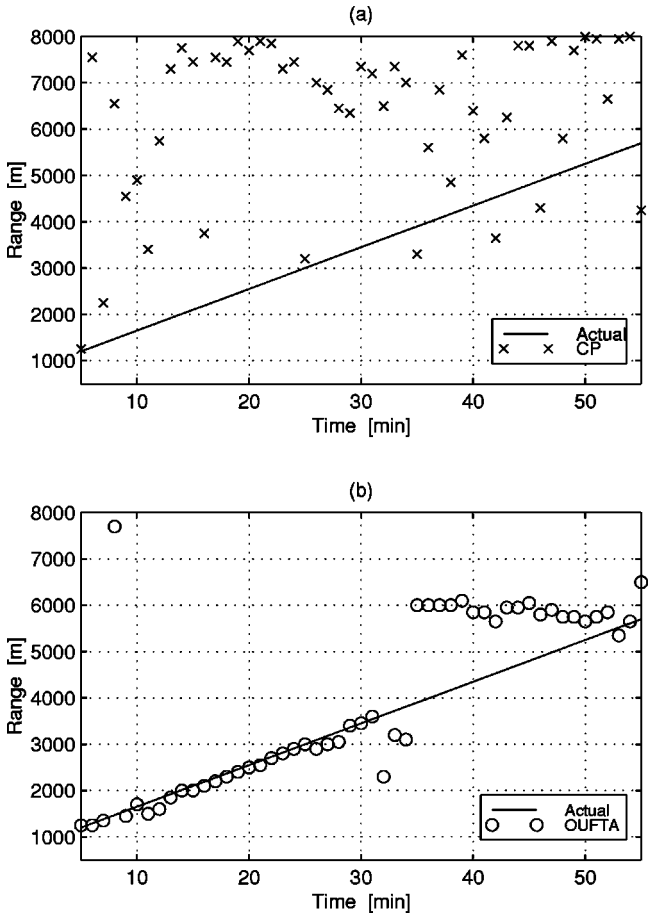


FIG. 4. SWellEX-1 experiment narrowband (195 Hz) range tracking performance for: (a) conventional processing, (b) OUFTA.

applied to these data for a single frequency of 195 Hz. This narrowband frequency was chosen in order to have data which is too noisy to successfully process independently, and to fully illustrate the benefits of implementing an optimal approach. The narrowband results are discussed in Sec. V B. The wideband and narrowband results should not be directly compared since the available SNR is not the same for the two cases.

A. Wideband results using TBD beamforming

Results attained by applying conventional and TBD beamforming methods to wideband data consisting of four narrowband tones at frequencies of 70, 95, 145, and 195 Hz are presented in Figs. 2 and 3. The results are shown for the Bartlett processor, the MV beamformer with environmental perturbation constraints (MV-EPC),[6] and the MV beamformer with neighborhood location constraints (MV-NLC). The ambiguity surfaces are computed by first calculating the individual narrowband ambiguity surfaces and then forming the incoherent average across the four frequencies. The source location estimates are determined by choosing the maximum of the ambiguity surface over a search grid of 2–8 km in range and 0–127 m in depth. Estimates with errors greater than 800 m in range or 15 m in depth are not shown. The TBD processors assume equiprobable movement of the source within a ± 100 m by ± 5 m neighborhood between frames. Each data frame consists of sixteen data snapshots, and each ambiguity function is computed using eight data frames.

The range tracking performance of conventional methods and TBD methods is compared in Fig. 2, and the depth tracking performance is compared in Fig. 3. Note that there are several breaks in the range and depth track estimates when conventional processing methods are employed. In contrast, the TBD processing results show that a continuous track is maintained.

B. Narrowband results using Bayesian methods

The range and depth tracking results for conventional processing and the OUFTA applied to a single frequency (195 Hz) are presented in Figs. 4 and 5. In Figs. 4 and 5, the actual path taken by the source is represented by a solid line, the estimates shown for conventional processing (CP) are the individual source locations estimated by the OUFP at each observation, and the estimates shown for the OUFTA are the end points of the most likely path after each observation. The search grid extends from 0.5 to 8 km in range and 0 to 200 m in depth, with the size of each grid point defined to be 50 m in range by 2 m in depth. The environmental integration is performed using 100 independent random realizations of the environment. The transition matrix used by the OUFTA is uniform with a maximum source motion of ± 100 m in range

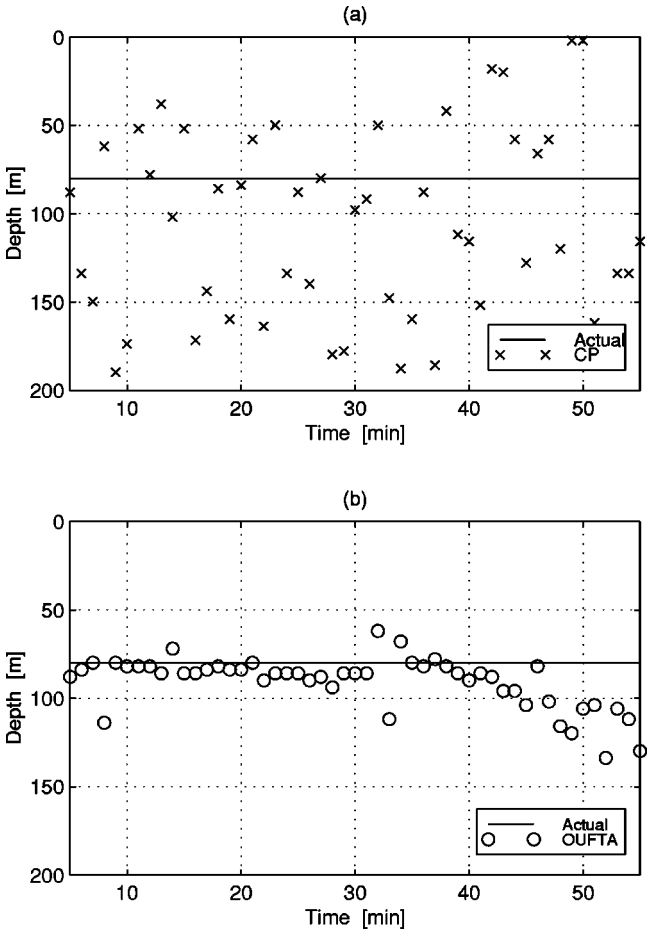


FIG. 5. SWellEX-1 experiment narrowband (195 Hz) depth tracking performance for: (a) conventional processing, (b) OUFTA.

and ± 4 m in depth, for a total of 25 possible transitions. The frequency domain data are sampled once every minute in order to obtain the observed data. Each ambiguity function is computed from a single data snapshot. A total of 50 min of data (51 observations) are processed, and the time scale is referenced to the beginning of the data collection.

By comparing the range tracking results (Fig. 4) and the depth tracking results (Fig. 5), one can see that the OUFTA provides marked improvement in narrowband tracking capability over conventional processing when the assumption of a single source is valid. There is an unidentified source of interference located near the surface which appears about 25 min into the data collection. As these results show, the OUFTA provides fairly accurate source track estimates before the interfering source is present. However, the problem of designing an optimal algorithm for tracking sources in the presence of interference was not addressed here and is left for future research.

VI. CONCLUSIONS

This paper has illustrated the benefits of implementing an optimal tracking approach over conventional methods through application of these techniques to data collected during the SWellEX-1 experiment. Further, the relationship between the OUFTA and MV-TBD beamformers has been examined, revealing how an optimal tracking philosophy may be used to guide the modification of existing beamforming techniques. Conventional methods typically produce source track estimates which contain discontinuities. In contrast, the optimal tracking methods presented here incorporate *a priori* knowledge of the source dynamics in addition to the underlying uncertainties in the ocean environment. The results using the SWellEX-1 data illustrate that using this additional information enables the tracking processors to estimate the source track more accurately.

ACKNOWLEDGMENT

Support for this work has been provided by the Office of Naval Research.

[1] H. P. Bucker, "Use of calculated sound fields and matched-field detection to locate sound sources in shallow water," J. Acoust. Soc. Am. **59**, 368–373 (1976).
[2] A. Tolstoy, "Sensitivity of matched field processing to sound-speed profile mismatch for vertical arrays in a deep water Pacific environment," J. Acoust. Soc. Am. **85**, 2394–2404 (1989).
[3] D. F. Gingras, "Methods of predicting sensitivity of matched-field processors to mismatch," J. Acoust. Soc. Am. **86**, 1940–1949 (1989).
[4] M. B. Porter and A. Tolstoy, "The matched field processing benchmark problems," J. Comput. Acoust. **2**, 161–185 (1994).
[5] A. M. Richardson and L. W. Nolte, "*A posteriori* probability source localization in an uncertain sound speed, deep ocean environment," J. Acoust. Soc. Am. **89**, 2280–2284 (1991).
[6] J. L. Krolik, "Matched-field minimum variance beamforming in a random ocean channel," J. Acoust. Soc. Am. **92**, 1408–1419 (1992).
[7] J. R. Daugherty and J. F. Lynch, "Surface wave, internal wave, and source motion effects on matched field processing in a shallow water waveguide," J. Acoust. Soc. Am. **87**, 2503–2526 (1990).
[8] H. Bucker, "Matched-field tracking in shallow water," J. Acoust. Soc. Am. **96**, 3809–3811 (1994).
[9] C. A. Zala and J. M. Ozard, "Matched-field processing for a moving source," J. Acoust. Soc. Am. **92**, 403–417 (1992).
[10] S. L. Tantum and L. W. Nolte, "Tracking and localizing a moving source in an uncertain shallow water environment," J. Acoust. Soc. Am. **103**, 362–373 (1998).
[11] R. T. Bachman, P. W. Schey, N. O. Booth, and F. J. Ryan, "Geoacoustic databases for matched-field processing: Preliminary results in shallow water off San Diego, California," J. Acoust. Soc. Am. **99**, 2077–2085 (1996).
[12] K. Harmanci, J. Tabrikian, and J. L. Krolik, "Relationships between adaptive minimum variance beamforming and optimal source localization," IEEE Trans. Signal Process. **48**, 1–12 (2000).
[13] A. Steinhardt, "Adaptive multisensor detection and estimation," in *Adaptive Radar Detection and Estimation*, edited by Simon Haykin and Allan Steinhardt (Wiley, New York, 1992).
[14] J. A. Shorey, L. W. Nolte, and J. L. Krolik, "Computationally efficient Monte Carlo estimation algorithms for matched field processing in uncertain ocean environments," J. Comput. Acoust. **2**, 285–314 (1994).
[15] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," IEEE Trans. Inf. Theory **IT-13**, 260–269 (1967).
[16] G. D. Forney, Jr., "The Viterbi algorithm," Proc. IEEE **61**, 268–278 (1973).
[17] H. Schmidt, A. B. Baggeroer, W. A. Kuperman, and E. K. Sheer, "Environmentally tolerant beamforming for high-resolution matched field processing: Deterministic mismatch," J. Acoust. Soc. Am. **88**, 1802–1810 (1990).
[18] K. Harmanci, J. Tabrikian, and J. L. Krolik, "Maximum likelihood track-before-detect matched-field beamforming with SWellEx data," J. Acoust. Soc. Am. **103**, 2898 (1998).
[19] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP Press, Woodbury, NY, 1994).
[20] M. B. Porter and E. L. Reiss, "A numerical method for ocean acoustic normal modes," J. Acoust. Soc. Am. **76**, 244–252 (1984).
[21] M. B. Porter, "The KRAKEN normal mode program," Report No. SM-245, SACLANT Undersea Research Centre, La Spezia, Italy, 1991.

The optimal stack spacing for thermoacoustic refrigeration

M. E. H. Tijani,[a] J. C. H. Zeegers, and A. T. A. M. de Waele
Department of Applied Physics, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

(Received 28 November 2001; revised 19 April 2002; accepted 25 April 2002)

The characteristic pore dimension in the stack is an important parameter in the design of thermoacoustic refrigerators. A quantitative experimental investigation into the effect of the pore dimensions on the performance of thermoacoustic devices is reported. Parallel-plate stacks with a plate spacing varying between 0.15 and 0.7 mm are manufactured and measured. The performance measurements show that a plate spacing in the stack of 0.25 mm ($2.5\delta_k$) is optimum for the cooling power. A spacing of 0.4 mm ($4\delta_k$) leads to the lowest temperature. The optimum spacing for the performance is about 0.3 mm ($3\delta_k$). It is concluded that a plate spacing in the stack of about three times the penetration depth should be optimal ($3\delta_k$) for thermoacoustic refrigeration. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1487842]

PACS numbers: 43.35.Ud, 43.35.Ty [RR]

I. INTRODUCTION

Thermoacoustic refrigerators are devices which use sound to generate cooling power. They consist mainly of an acoustic resonator filled with a gas. In the resonator a stack consisting of a number of parallel plates and two heat exchangers are installed, as shown in Fig. 1. A loudspeaker generates a standing wave in the resonance tube. The thermal interaction of the oscillating gas with the surface of the stack generates a heat transfer from one end of the stack to the other end. The heat exchangers are necessary to exchange heat with the surroundings, at the cold and hot sides of the stack. The way these systems work is explained extensively in the literature.[1,2]

The stack forms the heart of the refrigerator where the heat-pumping process takes place, and it is thus an important element which determines the performance of the refrigerator. A detailed experimental study of the effect of the plate spacing in the stack on the performance of thermoacoustic systems is still lacking in the literature. In this paper the result of a systematic investigation into the effect of the spacing in the stack on the behavior of the refrigerator is reported. Stacks with plate spacing varying between 0.15 and 0.7 mm have been constructed and measured. The manufacturing technique along with the measurement results are described in the following.

II. STACK MATERIAL AND GEOMETRY

The heat conductivity through the stack material and gas in the stack region has a negative effect on the performance of thermoacoustic refrigerators.[1,3] The stack material must have a low thermal conductivity and a heat capacity much larger than the heat capacity of the working gas in order that the temperature of the stack plates remains steady. The material Mylar is chosen, as it has a low heat conductivity (0.16 W/mK) and is proposed in thicknesses of 10–500 μm.

There are many geometries which the stack can have: parallel plates, circular pores, pin arrays, triangular pores, etc. The geometry of the stack is expressed in Rott's function f_k.[4] As can be seen from the energy equation, the cooling power is proportional to $\mathrm{Im}(-f_k)$.[4,5] Figure 2 shows the real and imaginary parts of f_k for some geometries as functions of the ratio the hydraulic radius r_h and the thermal penetration depth. The pin arrays and parallel-plate stacks are the most effective geometries. Since the pin-array stack is too difficult to manufacture, it is decided to use a stack made of parallel plates. Figure 3 shows an illustration of a parallel-plate stack. We note that for parallel-plate stack $r_h = y_0$.

The selection of an operating frequency of 400 Hz, an average pressure of 10 bar, and helium as working gas, determines the thermal and viscous penetration depths given by, respectively,

$$\delta_k = \sqrt{\frac{2K}{\rho c_p \omega}}, \tag{1}$$

and

$$\delta_\nu = \sqrt{\frac{2\mu}{\rho\omega}}, \tag{2}$$

where K is the thermal conductivity of the gas, μ is the viscosity of the gas, ρ is the density of the gas, c_p is the isobaric specific heat per unit mass, and ω is the angular frequency of the sound wave. Using Eqs. (1) and (2) for our system we have $\delta_k = 1$ mm and $\delta_\nu = 0.08$ mm.

As can be seen from Fig. 2 for a parallel-plate stack $\mathrm{Im}(-f_k)$ has a maximum for $y_0/\delta_k = 1.1$. Since the spacing in the stack is $2y_0$, this means that the theoretical optimal spacing is about 0.22 mm ($2.2\delta_k$). Using an analogous analysis, Arnott *et al.*[5] showed an optimal spacing of about 0.26 mm. In order not to alter the acoustic field, it was stated[2] to use a spacing of $2\delta_k$ to $4\delta_k$. It is decided to investigate experimentally the effect of the spacing in the stack by constructing parallel-plate stacks with spacings varying between 0.15 and 0.7 mm.

[a]Electronic mail: m.e.h.tijani@tue.nl

 0001-4966/2002/112(1)/128/6/$19.00

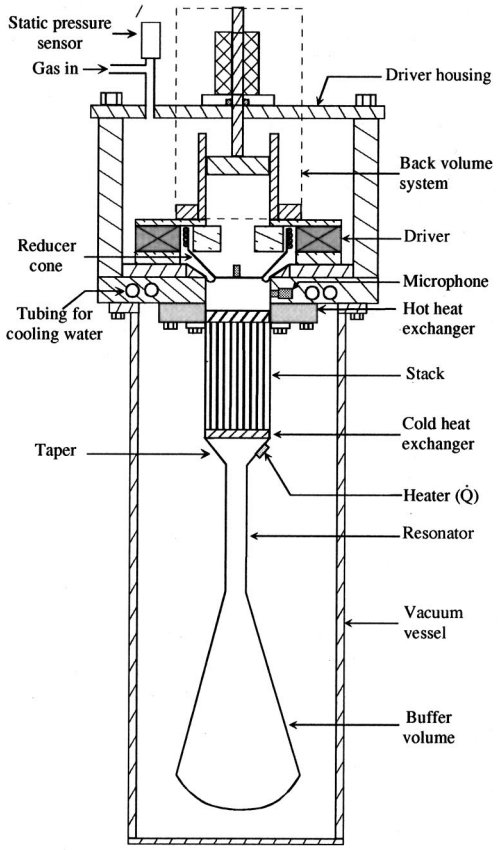


FIG. 1. Cross-sectional illustration of the thermoacoustic refrigerator, showing the different parts.

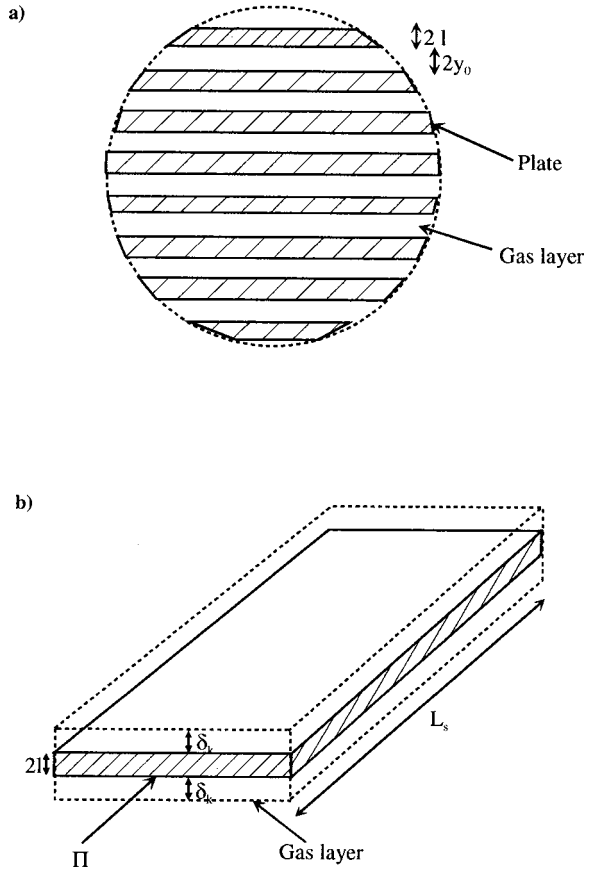


FIG. 3. Illustration of a parallel-plate stack. (a) Cross-sectional view of the stack, the plates are $2l$ thick and are $2y_0$ spaced. (b) A plate of the stack is shown.

III. MANUFACTURING OF THE PARALLEL-PLATE STACKS

The parallel-plate stacks consist of parallel plates which are spaced by fishing line spacers glued between the plates, as shown in Fig. 4(b). This type of stack is more difficult to manufacture than the spiral stack which one rolled up, results in a rigid structure.[6,7] In parallel-plate stacks the adhesion between the plates is ensured by the spacers which provide only a small adhesion surface. This leads to a fragile structure. One of the manufacturing problems to overcome was the gluing of nylon fishing line onto Mylar plates. Good results are obtained with 3M Scotch-Grip 7312 glue.[3]

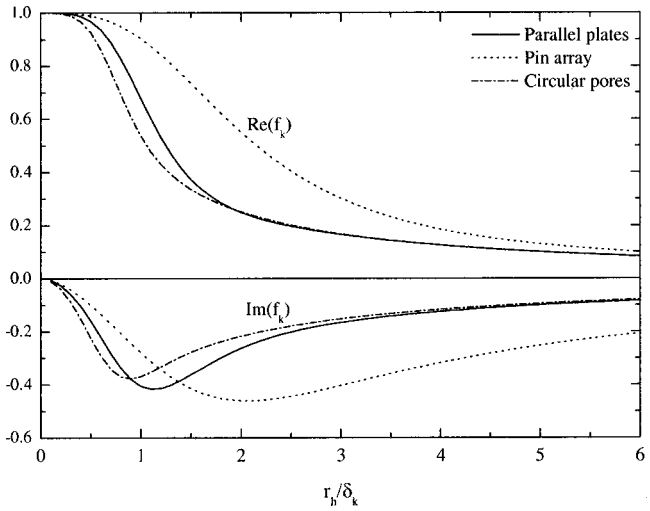


FIG. 2. Imaginary and real parts of the Rott function f_k as function of the ratio of the hydraulic radius and the thermal penetration depth. Three geometries are considered. The pin arrays and parallel-plate stacks are the best. The pin arrays an internal radius $r_i = 3\delta_k$ is used in the calculations.

For ease and speed of construction a special mounting setup was built, as shown in Fig. 4(a). The setup consists of a work plate which supports vertical rods. The diameter of the rods can be varied, depending on the spacing between the plates. Since fishing lines with different diameters are used to realize different stacks with different spacings, the number of rods is increased as the diameter becomes smaller. This increase in the number of rods is needed to obtain a uniform channel structure and enough surface for adhesion.

A fishing reel is mounted on a vertical bar. This reel is used to provide the needed tension on the fishing line during the winding process. A plastic syringe is used as a reservoir for glue as well as handle to stretch the line. The manufacturing procedure of the parallel plate stacks is as follows. The 0.1-mm-thick Mylar plates are first cut to the needed dimensions. A plate is placed between the brass rods as shown in Fig. 4(a). The fishing line enters the syringe from the top hole, through the glue and exits the syringe via the bottom hole. The latter has a little oversized diameter so that the thickness of the film of glue on the fishing line is 0.15 mm. In this way a uniform film of glue is obtained on the line. The fishing line is first wound around the first aluminum rod, and then around the brass rods [Fig. 4(a)]. At the end of

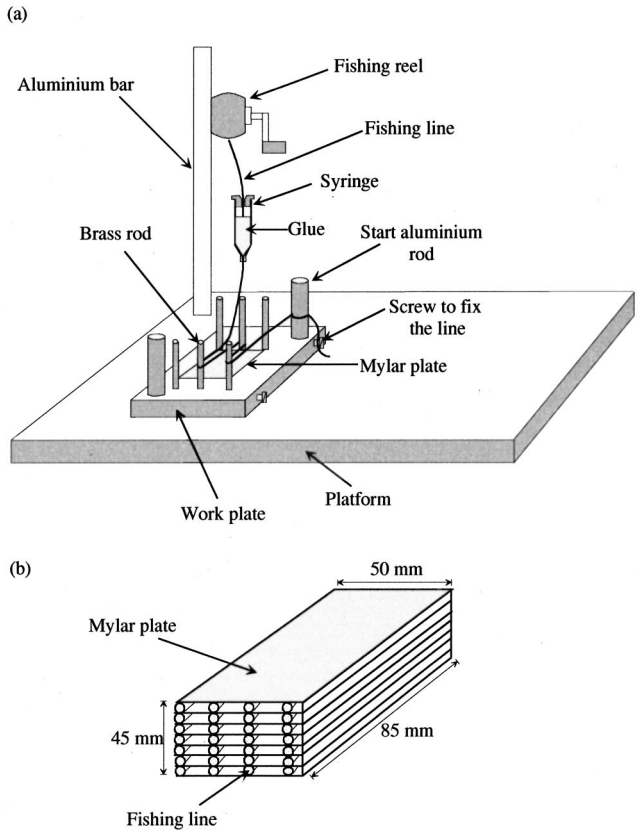


FIG. 4. Illustration of the manufacturing technique of the parallel-plate stack. A plate is placed between the brass rods, then the glue coated fishing line coming from the syringe is would around the rods to get parallel strings. The strings are then pressed onto the plate and another plate may be placed. The procedure is repeated until the needed height of the stack is reached. (a) Setup. (b) A parallel-plate stack as released from the work plate.

the winding process the line is wound around the last aluminum rod and fixed by a screw to the edge of the work plate. A press (stamper) is used to affix the fishing lines to the plate. Then, a new plate is placed on the strings and the winding process starts again. This stacking procedure is repeated until the needed height of the stack is obtained. The stack is pressed by a stainless steel block and the glue is allowed to dry for at least one day. The stack is then shaped into a cylindrical form of diameter 38 mm, using a milling technique. The building of a stack takes one day, and the shaping process two days. We note that care was taken during the winding process so that just enough tension is used, otherwise the stack will be deformed. In total seven stacks have been successfully manufactured with a spacing varying between 0.15 and 0.7 mm. It should be noted that all of the stacks have the same diameter, the same length, the same plate thickness, and are placed at the same location in the resonator.

IV. EXPERIMENTAL SETUP

An illustration of the thermoacoustic refrigerator used for the measurements is shown in Fig. 1. A modified loudspeaker (driver) is attached to one end of an acoustic resonator which is filled with helium gas at 10 bar. A stack and two heat exchangers are placed in the resonator. The loudspeaker sustains an acoustic standing wave in the gas, at the fundamental resonance frequency of the resonator. The other end of the resonator terminates in a buffer volume which simulates an open end. The system is essentially a quarter-wavelength system, with a pressure antinode at the loudspeaker end and pressure node at the buffer volume end. The acoustic driver consists of a modified moving-coil loudspeaker, from which the fabric dome was cut off near the voice coil and replaced by a thin walled light aluminum cone glued onto the voice coil. A rolling diaphragm is used to seal the resonator from the driver housing, as shown in Fig. 1. A detailed description of the setup can be found elsewhere.[3,8,9]

The stack is mounted in a stack holder which is made out of POM-Ertacetal, a material with high rigidity and low thermal conductivity. The holder has an inner diameter of 38 mm, a wall thickness of 2 mm, and a length of 85 mm. It has two flanges for connections to the cold and hot heat exchanger flange. The holder is connected to the flange of the hot heat exchanger via six bolts. As this junction is at room temperature a rubber O-ring, mounted into the flange of the hot heat exchanger, is used for sealing. At the cold heat exchanger side, the stack holder is attached to the tapered part via twelve bolts. Because this side cools to below $-40\,^{\circ}\mathrm{C}$, we used an indium O-ring to seal this junction. Different stacks can be interchanged easily so that different performance measurements can be carried out.

The duty of a refrigerator is to remove a heat quantity ($\dot{Q}_C$) at a low temperature (T_C) and to supply a heat quantity ($\dot{Q}_H$) to the surroundings at a high temperature (T_H). To accomplish this process a net work input ($\dot{W}$) is required. The performance of the refrigerator, called the coefficient of performance (COP), is defined as

$$\mathrm{COP}=\frac{\dot{Q}_C}{\dot{W}}. \tag{3}$$

The Carnot coefficient of performance is defined as

$$\mathrm{COPC}=\frac{T_C}{T_H-T_C}. \tag{4}$$

This forms the maximum performance for all refrigerators, where T_H is the temperature of the hot heat exchanger, and T_C is the temperature of the cold heat exchanger. The coefficient of performance relative to Carnot's coefficient of performance is defined as

$$\mathrm{COPR}=\frac{\mathrm{COP}}{\mathrm{COPC}}. \tag{5}$$

The performance measurements for the refrigerator are presented in plots of COP, COPR, and ΔT given by

$$\Delta T=T_C-T_H, \tag{6}$$

as functions of the total heat load. To understand the behavior of the cooler as a function of the varied parameters, the measured ΔT data will be fit with

$$\Delta T=\Delta T_0+\alpha\dot{Q}, \tag{7}$$

where $\dot{Q}$ is the total heat load which is the sum of the heat applied by the heater to the cold heat exchanger and the

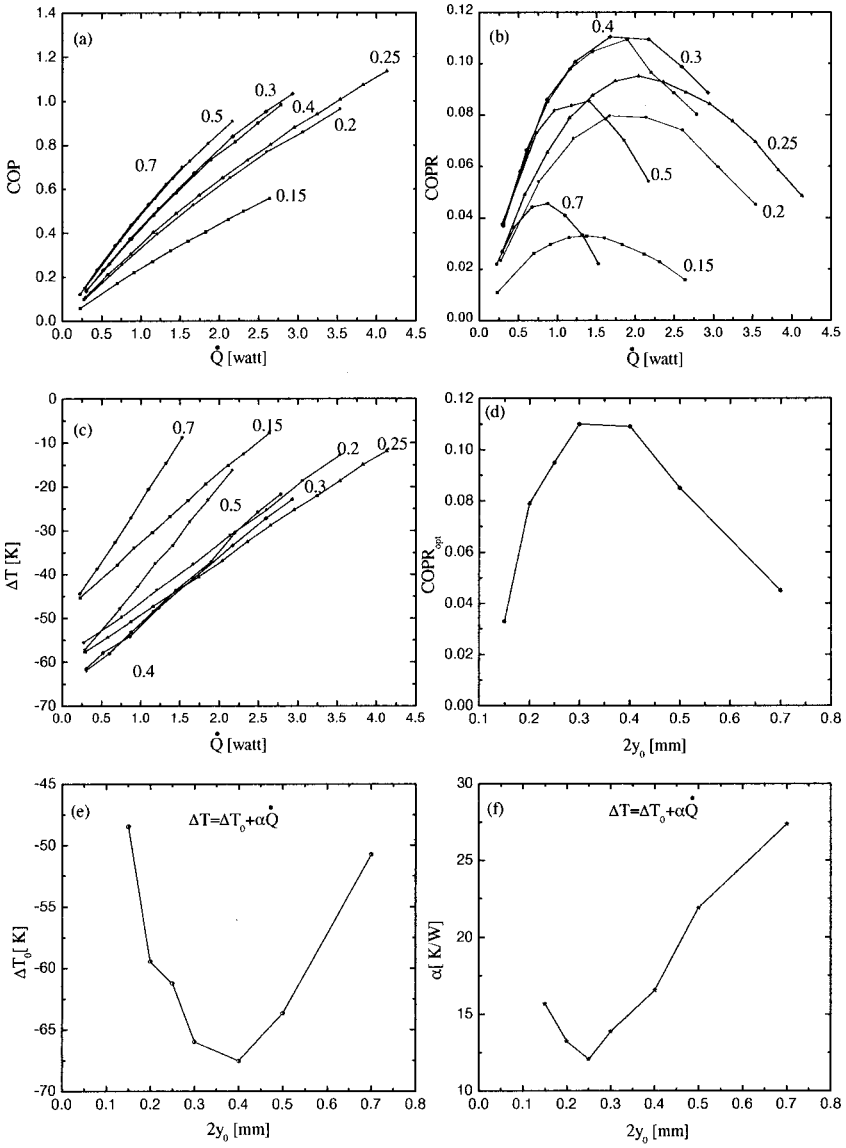


FIG. 5. Measurements with different stacks having plate spacing between 0.15 and 0.7 mm. The working gas is helium at 10 bar, and the drive ratio is 1.4%. In (a), (b), and (c) the values of the spacing $2y_0$ are indicated at the corresponding curves. (a) COP. (b) COPR. (c) ΔT. (d) Optimal COPR corresponding to the peak COPR as function of $2y_0$. (e) ΔT_0 as function of $2y_0$. (f) α as function of $2y_0$.

external heat leak.[3] During a given performance measurement the drive ratio, defined as the ratio of the dynamic pressure to the average pressure, is held constant while stepwise the heat load is increased and the temperature is allowed to stabilize. The heat load is applied by an electric heater which is placed at the cold heat exchanger. Two thermometers are used to monitor the temperature at the hot heat exchanger, and the cold heat exchanger. The thermometers are read by a multimeter which in turn is read by a computer. For all measurements and DELTAE[10] calculations reported in this paper, helium at an average pressure of 10 bar is used as working gas. A drive ratio $D=p_1/p_m=1.4\%$ is also used, where p_m is the average pressure and p_1 is the dynamic pressure.

V. RESULTS AND DISCUSSION

Performance measurements have been done with different stacks, having a plate spacing, $2y_0$, which varies between 0.15 and 0.7 mm. Figure 5 shows the measurements results. As can be seen from Fig. 5(a), the COP increases as function of the heat load for all stacks. Additionally, the slope of COP is an increasing function of $2y_0$. The COPR shows a parabolic behavior with a maximum [Fig. 5(b)] for all stacks, but the magnitude and the position of the peak changes as the spacing changes. To get a better insight into the effect of the spacing on COPR, the maxima from Fig. 5(b) are plotted as function of $2y_0$ in Fig. 5(d). The COPR shows a maximum at a spacing of about 0.3 mm ($3\delta_k$).

The temperature differences as functions of the heat load and the spacing are shown in Fig. 5(c). For all stacks, the dependence on the heat load is linear but again the slope changes as the spacing changes. For a better illustration of the influence of the spacing on the temperature difference, the curves in Fig. 5(c) are fitted with Eq. (7). Figure 5(e) shows that the temperature difference without heat load, ΔT_0, has an optimum around a plate spacing of about 0.4 mm ($4\delta_k$), which we did not expect. On the other hand, Fig. 5(f) shows that the coefficient α has an optimum at a plate spacing of about 0.25 mm ($2.5\delta_k$). Since a minimum for α leads to a maximum for the cooling power, the experimental optimal spacing is in agreement with the theoretical expectation.

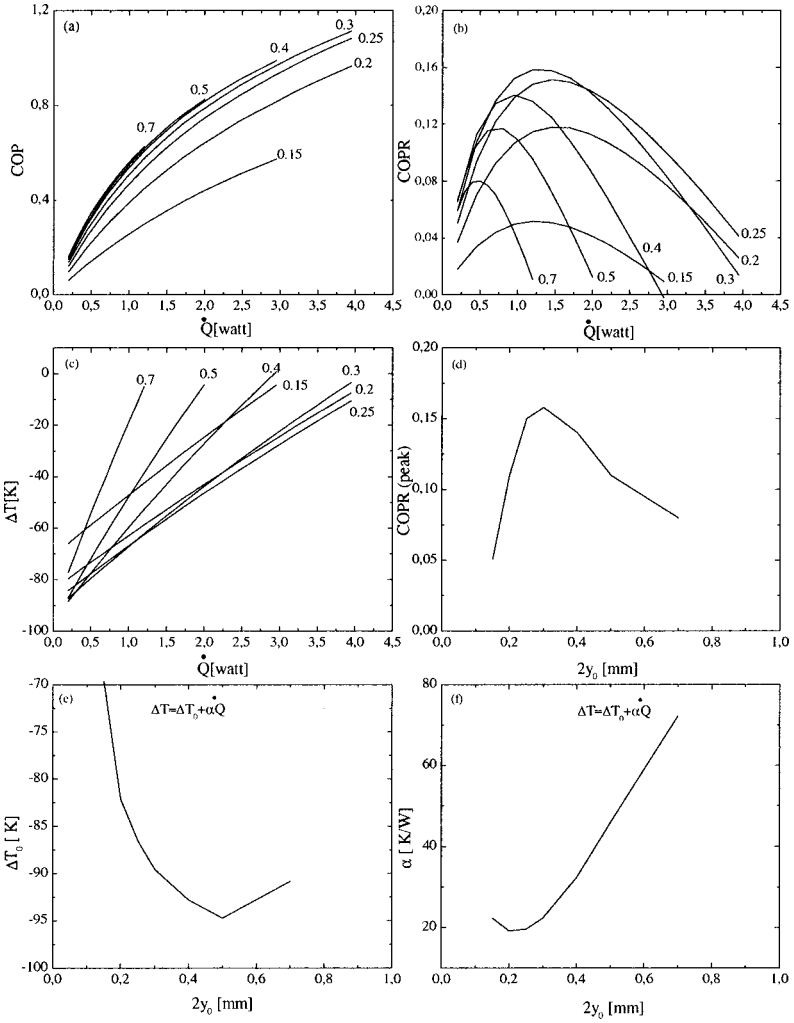


FIG. 6. DELTAE calculations with different stacks having plate spacing between 0.15 and 0.7 mm. The working gas is helium at 10 bar and the drive ratio is 1.4%.

The behavior of the cooling power (α) and COPR can be explained as follows: For plate spacings larger than about $2\delta_k$, the boundary layer picture holds and the cooling power is proportional to the product of the total perimeter of the stack in the cross section (Π) and the thermal penetration depth δ_k[1] (Fig. 3). An increase in the spacing means a decrease of the number of plates and hence a decrease of the perimeter. This results in a decrease of the cooling power and COPR above around $3\delta_k$. On the other hand, once the spacing becomes smaller than about $2.5\delta_k$, the thermoacoustic shuttle effect will be reduced, which will decrease the cooling power. Additionally, the whole gas layer between the plates contributes to the viscous shear and the small spacing induces higher velocities and hence extra viscous losses. This explains the rapid decreasing of COPR for spacings smaller than about $3\delta_k$. Since for the stacks having smaller spacing more fishing lines are used to keep the plates parallel and to obtain a uniform channel structure, this has as effect to change the geometry from parallel plate toward rectangular channel, which is less efficient. This may also contribute to the rapid decrease of the performance below about $3\delta_k$. It should be noted that the porosity of the stack, i.e., the ratio of the gas area to the total area in the cross section, varies from 87% for the stack with a plate spacing of 0.7 mm to 60% for the stack with a plate spacing of 0.15 mm.

The above-discussed apply to the stack position and length used in our refrigerator which are obtained from the optimization procedure used to design our refrigerator.[3,8,9] The effect of the stack position on the above-discussed optima will be discussed in Sec. VI, which is concerned with DELTAE calculations.

VI. DELTAE CALCULATIONS

The simulation program DELTAE[10] has been used to predict the performance of the refrigerator. DELTAE is a computer program which resolves the thermoacoustic linear equations in a geometry given by the user. As it is a linear tool, the nonlinearities, which may occur in practice, are not included. The calculations for the different stacks are shown in Fig. 6. The general trend is in agreement with the measurements. Figure 6(b) shows the behavior of the COPR as a function of the heat load and for different stacks. Both the magnitude and position of the peak are influenced by the stack spacing. In Fig. 6(d), $\mathrm{COPR}_{\mathrm{opt}}$ is plotted versus the plate spacing in the stack. The $\mathrm{COPR}_{\mathrm{opt}}$ shows a maximum around 0.3 mm, which is in agreement with the measurements. The ΔT curves shown in Fig. 6(b) are fitted with Eq. (7). The coefficients ΔT_0 and α are plotted versus the plate spacing in Figs. 6(e) and (f), respectively. The coefficient α shows an optimum around 0.22 mm in agreement with the measurements [Fig. 6(f)]. The parameter ΔT_0 shows an op-

timum around 0.5 mm. It should be noted that the calculated ΔT curves are not exactly linear. This may explain the discrepancy between the experimental and calculated optima for ΔT_0. In summary, the calculations show a general trend which is in agreement with the measurements.

Because the experiments are lengthy and time consuming and seeing the good agreement between experiment and DELTAE calculations, we decided to simulate the effect of the stack position on the above-discussed optima using DELTAE. The calculations show that the optimal spacing for the cooling power (α) is nearly independent of the position of the stack in the acoustic field, and that the optimal spacing for COPR shifts to lower spacings as the position of the stack decreases (position zero corresponds to the driver location). This is due to the decrease of the velocity which results in a decrease of the viscous losses and hence an increase of the performance.

VII. CONCLUSION

The influence of the plate spacing in the stack on the behavior of the refrigerator was studied systematically. The performance of the thermoacoustic refrigerator, using stacks with plate spacings varying between 0.15 and 0.7 mm, is measured. We showed using a simplified analytical discussion that the parallel plate geometry is the best, and that a sheet spacing in the stack of about $2.5\delta_k$ should be optimal for the cooling power. As can be seen from Figs. 5(b) and (f), the cooling power for the stacks with a spacing of $2.5\delta_k$ and $3\delta_k$ are nearly equal. Additionally, the stack with the spacing of $3\delta_k$ has the highest performance. Taking into account this remark it is safe to say that the practical optimal spacing in the stack is about $3\delta_k$. This result is also confirmed by DELTAE calculations. The general trend of the DELTAE calculations show a satisfactory agreement with experiment.

ACKNOWLEDGMENTS

The authors would like to thank Greg Swift and Chris Espinoza of Los Alamos National Laboratories for their advice in the engineering of the parallel-plate stacks.

[1] G. W. Swift, ''Thermoacoustic engines,'' J. Acoust. Soc. Am. **84**, 1146–1180 (1988).
[2] J. C. Wheatley, T. Hofler, G. W. Swift, and A. Migliori, ''An intrinsically irreversible thermoacoustic heat engine,'' J. Acoust. Soc. Am. **74**, 153–170 (1983); ''Experiments with an intrinsically irreversible thermoacoustic heat engine,'' Phys. Rev. Lett. **50**, 499–502 (1983).
[3] M. E. H. Tijani, ''Loudspeaker-driven thermo-acoustic refrigeration,'' Ph.D. thesis, unpublished, Eindhoven University of Technology, 2001.
[4] G. W. Swift, ''Thermoacoustic engines and refrigerators,'' *Encyclopedia of Applied Physics*, edited by G. L. Trigg (Wiley-VCH, Verlag, Berlin, 1997), Vol. 21, pp. 245–264.
[5] W. P. Arnott, H. E. Bass, and R. Raspet, ''General formulation of thermoacoustics for stacks having arbitrarily shaped pore cross sections,'' J. Acoust. Soc. Am. **90**, 3228–3237 (1991).
[6] S. L. Garrett, J. A. Adeff, and T. J. Hofler, ''Thermoacoustic refrigerator for space applications,'' J. Thermophys. Heat Transfer **7**, 595–599 (1993).
[7] T. J. Hofler, ''Thermalization refrigerator design and performance,'' Ph.D. dissertation, Physics Department, University of California at San Diego, 1986.
[8] M. E. H. Tijani, J. C. H. Zeegers, and A. T. A. M. de Waele, ''Design of thermoacoustic refrigerators,'' Cryogenics **42**, 49–57 (2002).
[9] M. E. H. Tijani, J. C. H. Zeegers, and A. T. A. M. de Waele, ''Construction and performance of a thermoacoustic refrigerator,'' Cryogenics **43**, 59–66 (2002).
[10] W. C. Ward and G. W. Swift, ''Design environment for low-amplitude thermoacoustic engines,'' J. Acoust. Soc. Am. **95**, 3671–3672 (1994).

Prandtl number and thermoacoustic refrigerators

M. E. H. Tijani, J. C. H. Zeegers, and A. T. A. M. de Waele
Department of Applied Physics, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

(Received 28 November 2001; revised 25 April 2002; accepted 4 May 2002)

From kinetic gas theory, it is known that the Prandtl number for hard-sphere monatomic gases is 2/3. Lower values can be realized using gas mixtures of heavy and light monatomic gases. Prandtl numbers varying between 0.2 and 0.67 are obtained by using gas mixtures of helium–argon, helium–krypton, and helium–xenon. This paper presents the results of an experimental investigation into the effect of Prandtl number on the performance of a thermoacoustic refrigerator using gas mixtures. The measurements show that the performance of the refrigerator improves as the Prandtl number decreases. The lowest Prandtl number of 0.2, obtained with a mixture containing 30% xenon, leads to a coefficient of performance relative to Carnot which is 70% higher than with pure helium. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1489451]

PACS numbers: 43.35.Ud, 43.35.Ty [RR]

I. INTRODUCTION

The basic understanding of the physical principles underlying the thermoacoustic effect is well established and has been discussed in many papers.[1,2] However, a quantitative experimental investigation of the effect of some important parameters on the behavior of the thermoacoustic devices is still lacking. One of these parameters is the Prandtl number, a dimensionless parameter characterizing the ratio of kinematic viscosity to thermal diffusivity. Viscous friction has a negative effect on the performance of thermoacoustic systems.[2,3] Decreasing the Prandtl number generally increases the performance of thermoacoustic devices. Kinetic gas theory has shown that the Prandtl number for hard-sphere monatomic gases is 2/3. Lower Prandtl numbers can be realized using mixtures of heavy and light monatomic gases, for example, binary gas mixtures of helium and other noble gases.[3–5]

The calculations of the Prandtl number shown in this paper use more rigorous expressions for the transport coefficients than those used previously.[4,5] For all binary gas mixtures of helium with other noble gases, calculations show that the coefficient of performance of the refrigerator is maximized when the Prandtl number is near its minimum value.

This paper presents the results of an experimental investigation into the effect of Prandtl number on the performance of a thermoacoustic refrigerator using gas mixtures of helium with argon, krypton, and xenon. These combinations provided gas mixtures with Prandtl numbers varying between 0.2 and 0.67. These results are discussed below.

The thermoacoustic refrigerator used for the measurements is shown in Fig. 1. The acoustic resonator is filled with an inert gas at a pressure of 10 bar. A channeled stack is strategically located in the resonator to facilitate heat transfer. At both ends of the stack heat exchangers are installed. The temperature of the hot heat exchanger is fixed at room temperature by circulating water. At the cold heat exchanger cooling power is generated. A loudspeaker generates a standing wave in the resonance tube, causing the gas to oscillate while compressing and expanding. The interaction of the moving gas in the stack with the stack surface generates heat transport.[2] A detailed description of the refrigerator can be found in the literature.[3,6]

II. KINETIC THEORY OF BINARY GAS MIXTURES

The Prandtl number is given by

$$\sigma = \frac{\mu c_p}{K}, \tag{1}$$

where μ is the dynamic viscosity, K is thethermal conductivity, and c_p is the isobaric specific heat. The Prandtl number can also be written in terms of the thermal and viscous penetration depths δ_k and δ_ν as

$$\sigma = \left(\frac{\delta_\nu}{\delta_k}\right)^2, \tag{2}$$

where δ_k and δ_ν are given by

$$\delta_k = \sqrt{\frac{2K}{\rho c_p \omega}}, \tag{3}$$

and

$$\delta_\nu = \sqrt{\frac{2\mu}{\rho\omega}}. \tag{4}$$

Here, ω is the angular frequency of the sound wave and ρ is the density.

The viscosity has a negative effect on the performance of thermoacoustic devices, so a reduction of the effect of viscosity means an increase in efficiency. This can be accomplished by lowering the Prandtl number. In the Appendix a survey of the kinematic theory for pure and binary gas mixtures is given. The text of Hirschfelder *et al.*[7] forms the principal source for this survey. By substituting the expressions of viscosity μ_{mix}, thermal conductivity K_{mix}, and the isobaric specific heat c_p for binary gas mixtures into Eq. (1), the Prandtl number of binary gas mixtures can be calculated. The

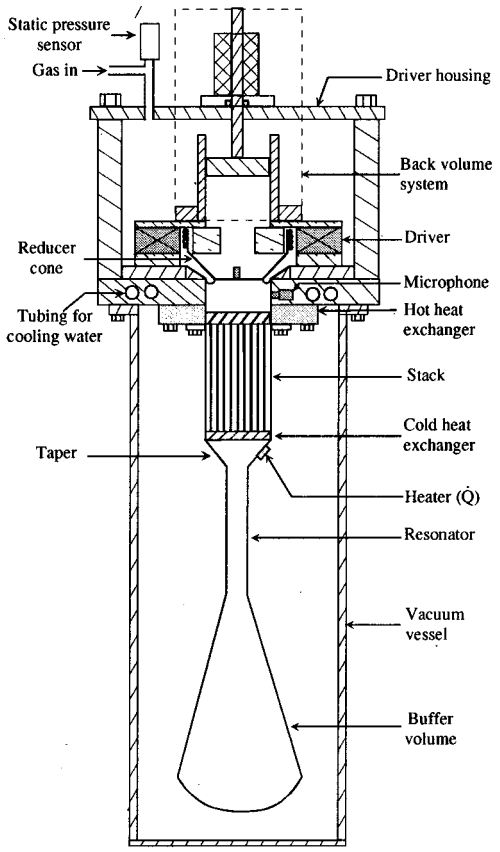


FIG. 1. Schematic diagram of the thermoacoustic refrigerator, showing the different parts.

resultant expression is too long to be given explicitly here. This was incorporated in a computer program to accomplish the calculations.

III. CALCULATION OF GAS MIXTURES PROPERTIES

In Fig. 2, some calculated properties of binary gas mixtures, consisting of He–Ne, He–Ar, He–Kr, and He–Xe, are plotted as functions of mole fraction, x, of the heavy component. The temperature and pressure used in the calculations are 250 K and 10 bar, respectively. These values apply to the experimental situation. Figure 2(a) shows that, for all mixtures, the density ρ increases linearly as a function of x, in accordance with Eq. (A14). From Figs. 2(b) and (c), it can be seen that the kinematic viscosity $\nu=\mu/\rho$ and thermal diffusivity $k=K/\rho c_p$ decrease as a function of x. This behavior is to be expected, since both properties are approximately inversely proportional to the square of the apparent mass of the mixture. But, in the range $0<x<0.4$ the rate of decrease for ν is larger than for k. This results in a decrease of the ratio of these two quantities which is the Prandtl number σ. For $x>0.4$, k still decreases while ν remains nearly constant; this results in an increase of σ until the value for a pure noble gas is reached. From this behavior of the Prandtl number decreasing and then increasing as function of the increase of the mole fraction of the heavier component, a minimum value is to be expected as shown in Fig. 2(f). Similar calculation results are obtained by Giacobbe[4] and Belcher *et al.*[5] using approximative expressions for the viscosity and thermal conductivity.

The behavior of the sound velocity a as a function of x is illustrated in Fig. 2(d). It is also a decreasing function of x, in accordance with Eq. (A16), since the apparent mass of the mixture increases. The behavior of c_p is shown in Fig. 2(e).

As can be seen from Fig. 2(f), the Prandtl number has a minimum at $x\approx 0.38$ for all binary gas mixtures. The value of σ at the minimum is a function of the molecular weight of the added heavier component. The lowest value of 0.2 is reached with the heaviest noble gas xenon. The minimum Prandtl number is plotted in Fig. 3 as a function of the molar mass of the heavy component. Extrapolation of the data for radon shows that using a helium–radon mixture, a Prandtl number of about 0.1 can be reached which can be considered as the lowest Prandtl number for helium–noble gases mixtures. The Prandtl number is also calculated for different gas mixtures, at different temperatures, as shown in Fig. 4. The influence of the temperature on the Prandtl number is small. The effect of temperature is most pronounced for He–Xe mixture. At the helium mole fraction corresponding to the minimum Prandtl number, a change in temperature of 100 K results in a relative change in the value of the Prandtl of only 6.5% for a helium–xenon mixture.

Figures 4(a) and (b) show the thermal and viscous penetration depths δ_k and δ_ν, respectively. The behavior of δ_ν as a function of x can be illustrated by means of Eq. (4). Since ν varies approximately as $1/\sqrt{M}$, it follows that δ_ν will vary as $1/\sqrt{4M}$, thus a decreasing behavior as shown in Fig. 4(b). A similar analysis can be done to explain the behavior of δ_k. It is interesting to note that δ_k has a maximum for a binary mixture at about $x=0.1$, as shown in Fig. 4(a). We note that for the calculations of δ_k and δ_ν, the wavelength is kept constant ($\lambda=2.35$ m) and the frequency is allowed to vary as the sound velocity varies with the composition.

The product ρa is calculated for a temperature of 250 K and at a pressure of 10 bar, and it is plotted in Fig. 4(c). The behavior can be understood as follows: the density increases linearly and a decreases as $1/\sqrt{M}$; as a result the product ρa increases as x increases, as $\sqrt{M}$. The product ρa will be needed later for discussions concerning the cooling power of the thermoacoustic refrigerator.

IV. PERFORMANCE CALCULATIONS

The purpose of a refrigerator is to remove a heat ($\dot{Q}_C$) at a low temperature (T_C) and to reject heat ($\dot{Q}_H$) to the surroundings at a high temperature (T_H). To accomplish this, work (W) is required. The coefficient of performance (COP) is defined as

$$\mathrm{COP}=\frac{\dot{Q}_C}{W}. \tag{5}$$

The quantity

$$\mathrm{COPC}=\frac{T_C}{T_H-T_C}, \tag{6}$$

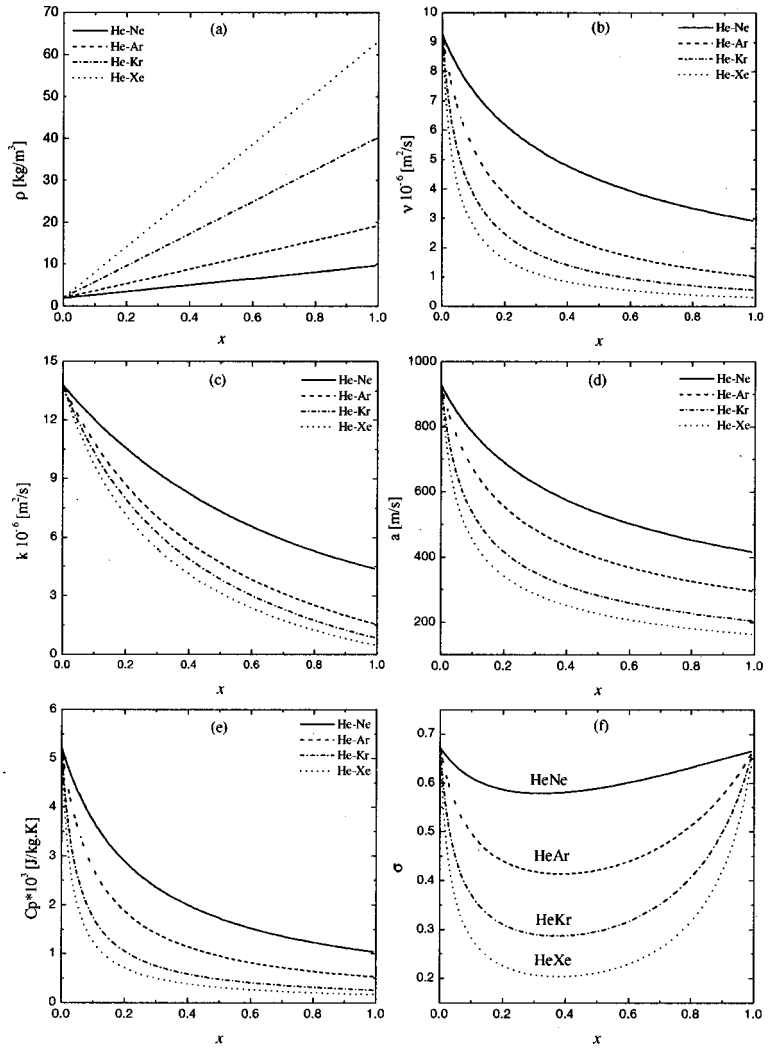


FIG. 2. Calculated properties of binary gas mixtures at a temperature of 250 K and a pressure of 10 bar. (a) Density. (b) Kinematic viscosity ν. (c) Thermal diffusivity k. (d) Speed of sound a. (e) The isobaric specific heat c_p. (f) Prandtl number σ.

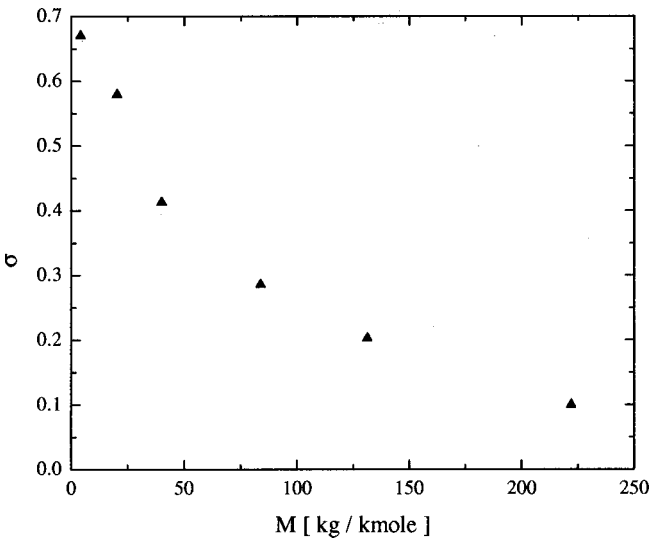


FIG. 3. Minimal Prandtl number, at a helium mole fraction of 0.62, as a function of the molecular weight of the heavy component. The first point corresponds to the Prandtl number of pure helium. The last point corresponds to the minimal Prandtl number of a helium–radon binary gas mixture obtained by extrapolation.

is called the Carnot coefficient of performance which defines the optimal performance for all refrigerators. The coefficient of performance relative to Carnot's coefficient of performance is defined as

$$\mathrm{COPR}=\frac{\mathrm{COP}}{\mathrm{COPC}}. \tag{7}$$

The performance measurements for the refrigerator are presented in plots of COP, COPR, and ΔT given by

$$\Delta T=T_{\mathrm{C}}-T_{\mathrm{H}}, \tag{8}$$

as functions of the total heat load, which is the sum of the heat load applied by the heater and the heat leak. To understand the behavior of the cooler as a function of the varied parameters, the measured ΔT data will be fit with the equation

$$\Delta T=\Delta T_0+\alpha\dot{Q}, \tag{9}$$

where $\dot{Q}$ is the total heat load.[3] During a given performance measurement, the drive ratio, defined as the ratio of the dynamic pressure to the average pressure, is kept constant

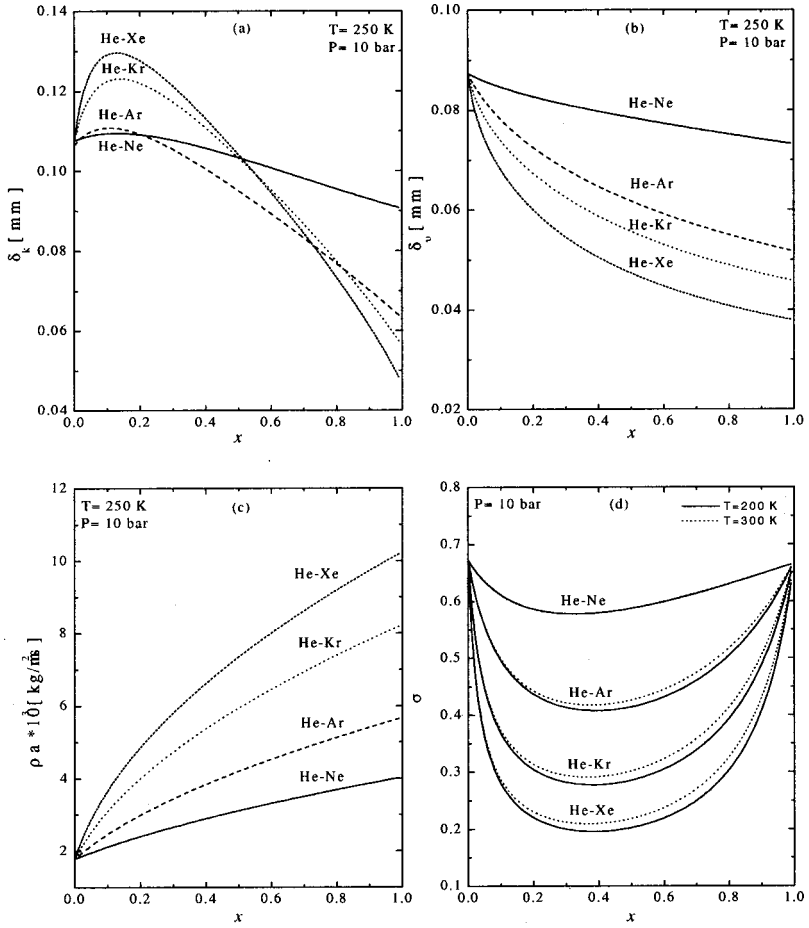


FIG. 4. The thermal and viscous penetration depths, δ_k and δ_ν, the product ρa, and Prandtl number σ for helium–other noble gas binary mixtures. We note that, for the calculations δ_k and δ_ν, the wavelength is kept constant ($\lambda = 2.35$ m) and the frequency is allowed to vary as the sound velocity varies with x. The temperature and pressure used in the calculations are also shown in the graphs. The Prandtl number has been calculated at two different temperatures.

while stepwise the heat load is applied to the cold heat exchanger and the temperature is allowed to stabilize. The heat load is applied by an electric heater placed at the cold heat exchanger. Two thermometers are used to monitor the temperatures at the hot heat exchanger and at the cold heat exchanger. Incorporating Eqs. (A5) and (A10) into the expressions for the energy flow and work flow in the stack,[8] the COP has been calculated as function of x. Figure 5 shows the behavior of COP for helium–xenon, helium–krypton, and helium–argon mixtures. For all binary gas mixtures, as the Prandtl number decreases the performance improves. A maximum is reached nearly at the point where the Prandtl number is a minimum (cf. Fig. 2).

V. MIXTURES PREPARATION

A container is used to prepare the gas mixtures. The low mole fraction component is first filled in the container until the pressure fraction needed is reached. Then, the second component is filled up to 33 bar. The mixture is allowed to reach equilibrium prior to filling the refrigerator to a pressure of 10 bar. The composition of the mixture was checked by measuring the resonance frequency of the refrigerator. This frequency is then related to the resonance frequency of the system for pure helium by the expression

$$\frac{f_{\mathrm{He}}}{f_{\mathrm{mix}}} = \sqrt{\frac{xM_2 + (1-x)M_{\mathrm{He}}}{M_{\mathrm{He}}}}, \tag{10}$$

where f_{He}, and f_{mix} are the resonance frequencies for pure helium and mixture, respectively. M_2 and x are the molar mass and mole fraction of the heavy component, and M_{He} is the molar mass of helium.

VI. MEASUREMENTS

With our thermoacoustic refrigerator[3,6] three different mixtures have been investigated: He–Ar, He–Kr, and He–Xe. The performance measurements were all made using a

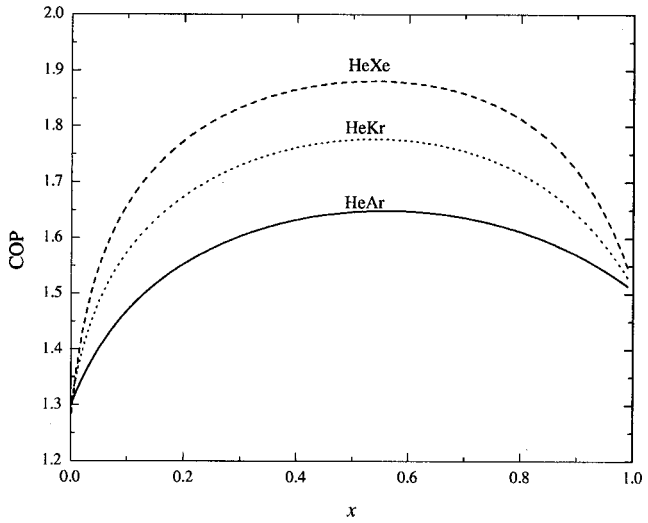


FIG. 5. Calculated COP as function of x.

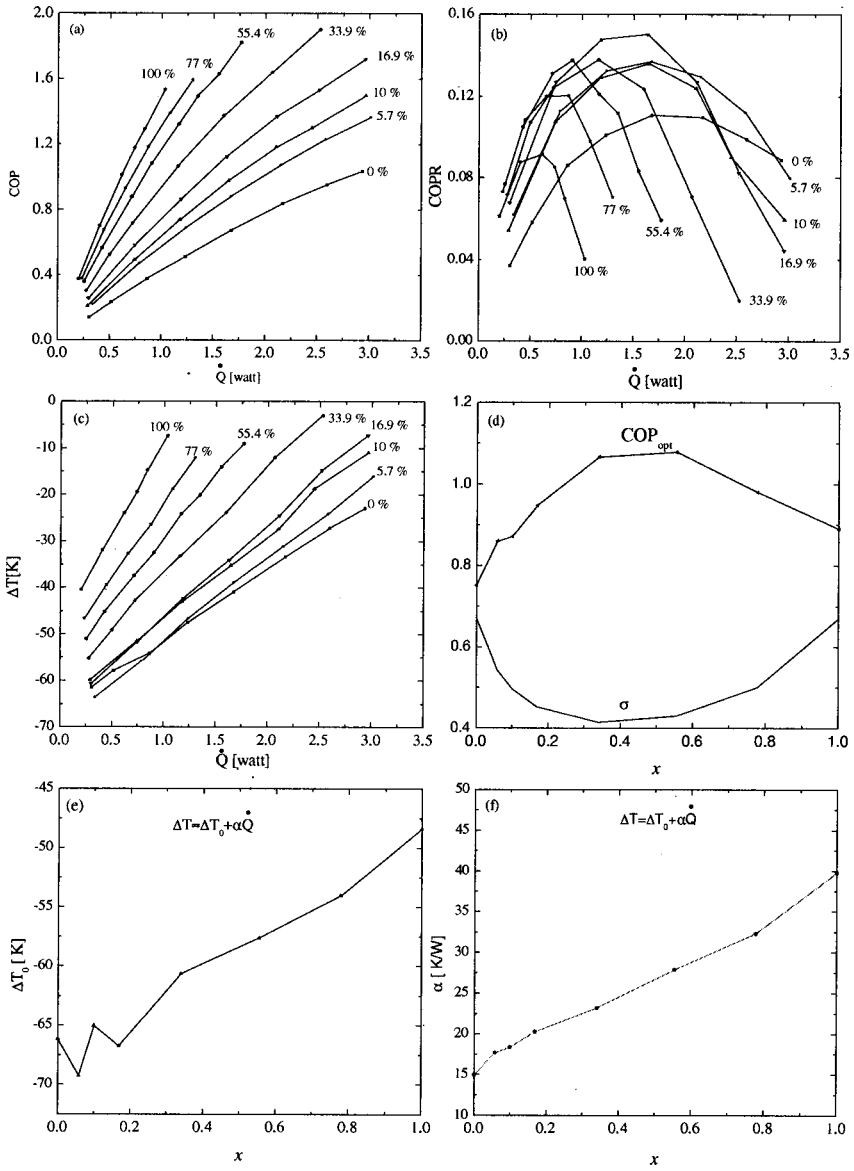


FIG. 6. Measurements with helium–argon mixtures. The average pressure is 10 bar, the drive ratio is 1.4%, and a stack with a plate spacing of 0.3 mm is used. In (a), (b), and (c) the values of x are indicated at the corresponding curves. (a) COP. (b) COPR. (c) ΔT. (d) Optimal COP corresponding to the peak COPR as function of the mole fraction argon x. (e) ΔT_0 as function of x. (f) α as function of x.

parallel-plate stack with spacing 0.3 mm, an average pressure of $p_m = 10$ bar, and a constant drive ratio of $D = p_1/p_m = 1.4\%$. The performance measurements for the refrigerator are presented in plots of COP, COPR, and ΔT as functions of the total heat load $\dot{Q}$.[3] During a given performance measurement, the drive ratio is held constant while stepwise the heat load is applied to the cold heat exchanger by a heater and the temperature is allowed to stabilize.[3]

A. Helium–argon

Binary helium–argon mixtures have been used with an argon mole fraction, x, varying between zero and 100%. The results are shown in Fig. 6. The COP increases as function of the heat load, as can be seen from Fig. 6(a). The slope of COP is an increasing function of x. From the calculations (Fig. 5) we expect that the COP will increase as a function of x, reach a maximum around 40%, and then decrease. This behavior is difficult to see from Fig. 6(a), because the increase of the COP slope is a consequence of two effects: the Prandtl number decrease and the decrease of the cooling power. The energy and work flux equations[8] show that the cooling power is proportional to the inverse of the product of the density times the speed of sound if the length of the resonator is held constant [Fig. 4(c)]. This factor, which is also proportional to the speed of sound in the mixture, decreases as x increases. Therefore, the cooling power decreases as x increases. The COPR as function of the heat load is shown in Fig. 6(b). For all mixtures the COPR shows a parabolic behavior with a maximum.

In Fig. 6(c), the temperature differences as functions of the heat load are shown. For all gas mixtures, the dependence is linear. The slope increases as x increases. In Fig. 6(d), the $\mathrm{COP}_{\mathrm{opt}}$ corresponding to the peak COPR (operation point) is plotted as function of x. This plot shows that the $\mathrm{COP}_{\mathrm{opt}}$ has the expected trend dependent on the Prandtl number.

To understand the effect of the Prandtl number on ΔT, the temperature curves of Fig. 6(c) are fitted with Eq. (9). The parameters ΔT_0 and α are plotted as functions of x in Figs. 6(e) and (f), respectively. As can be seen, ΔT_0 increases with x, with the largest temperature difference obtained with pure helium. The slope α also increases with x.

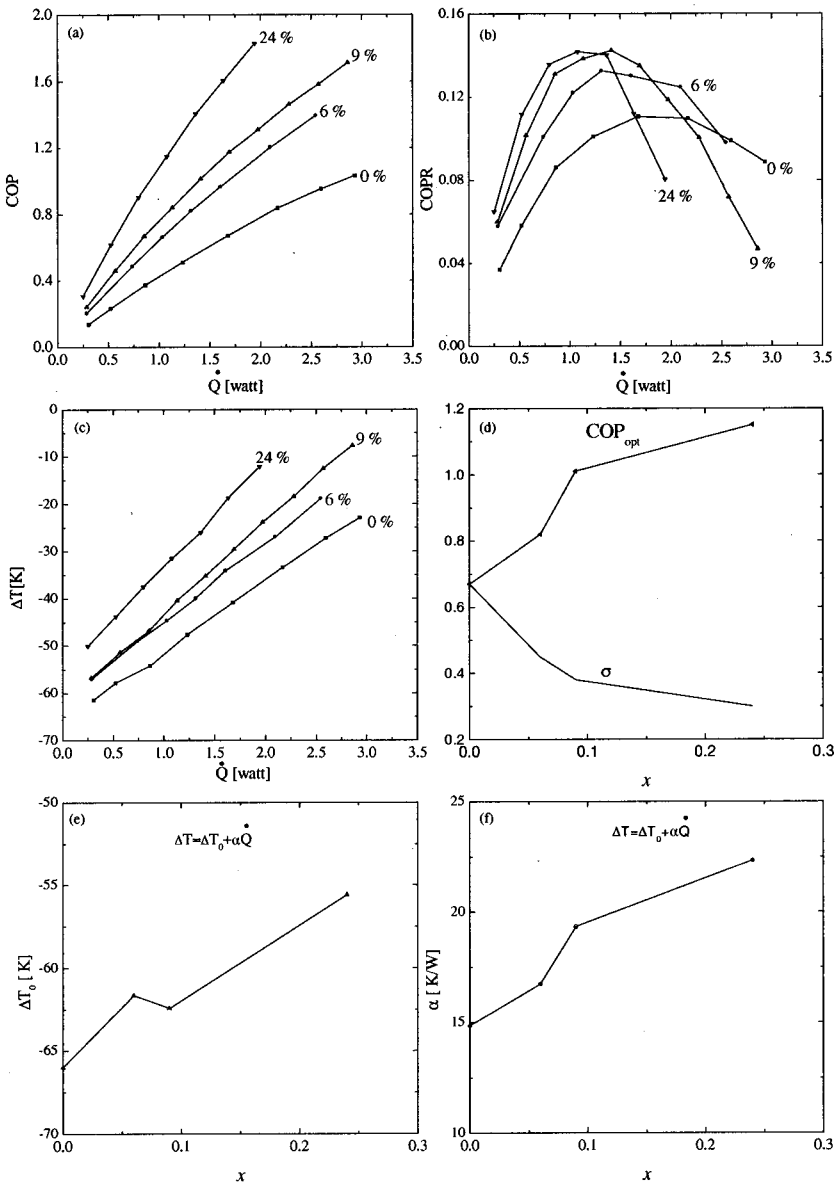


FIG. 7. Measurements with helium–krypton mixtures. The average pressure is 10 bar, the drive ratio is 1.4%, and a stack with a plate spacing of 0.3 mm is used. In (a), (b), and (c) the values of x are indicated at the corresponding curves. (a) COP. (b) COPR. (c) ΔT. (d) Optimal COP corresponding to the peak COPR as function of the mole fraction krypton x. (e) ΔT_0 as function of x. (f) α as function of x.

This means that the cooling power is a decreasing function of x. This is due to the increase of the density which increases as x increases.

In summary, the $\mathrm{COP}_{\mathrm{opt}}$ has the reverse behavior of the Prandtl number as function of x. The $\mathrm{COP}_{\mathrm{opt}}$ has a maximum when the Prandtl number has a minimum, since the viscous losses are minimum [Fig. 6(d)]. But, at the same time the addition of argon has a negative effect on the cooling power of the refrigerator, as the addition of argon increases the density which makes the cooling power decrease. The maximum COPR reached with helium–argon mixtures is 14.5% with a mixture containing 17% argon. This is a 40% improvement in comparisons with pure helium.

B. Helium–krypton

Measurements with helium–krypton mixtures are shown in Fig. 7. As can be seen from Fig. 7(d), the $\mathrm{COP}_{\mathrm{opt}}$ increases as the Prandtl number σ decreases. The calculated Prandtl number for the different mixtures is also plotted in this figure. The ΔT curves from Fig. 7(c) are fitted with Eq. (9) and the determined parameters ΔT_0 and α are plotted in Figs. 7(e) and (f). These parameters increase as functions of x, which means that the cooling power decreases. A maximal COPR of 14.3% is obtained with $x=24\%$, which has the lowest Prandtl number. The same general behavior as for helium–argon can be seen, except that the slopes of ΔT_0 and α are steeper in the case of helium–krypton. This means that the cooling power of the refrigerator is more heavily influenced by the addition of krypton than by the addition of argon. This is due to the higher density of krypton.

C. Helium–xenon

Figure 8 shows the results using helium–xenon mixtures. The general behavior of the COP and COPR is similar to those for He–Ar and He–Kr. As can be seen from Fig. 8(b), the magnitude of the maximum of COPR increases as the mole fraction xenon x increases and reaches a maximum of 17% for $x=0.3$. Figure 8(d) shows that the $\mathrm{COP}_{\mathrm{opt}}$ increases when σ decreases. The parameters ΔT_0 and α are plotted in Figs. 8(e) and (f). Again, these parameters increase

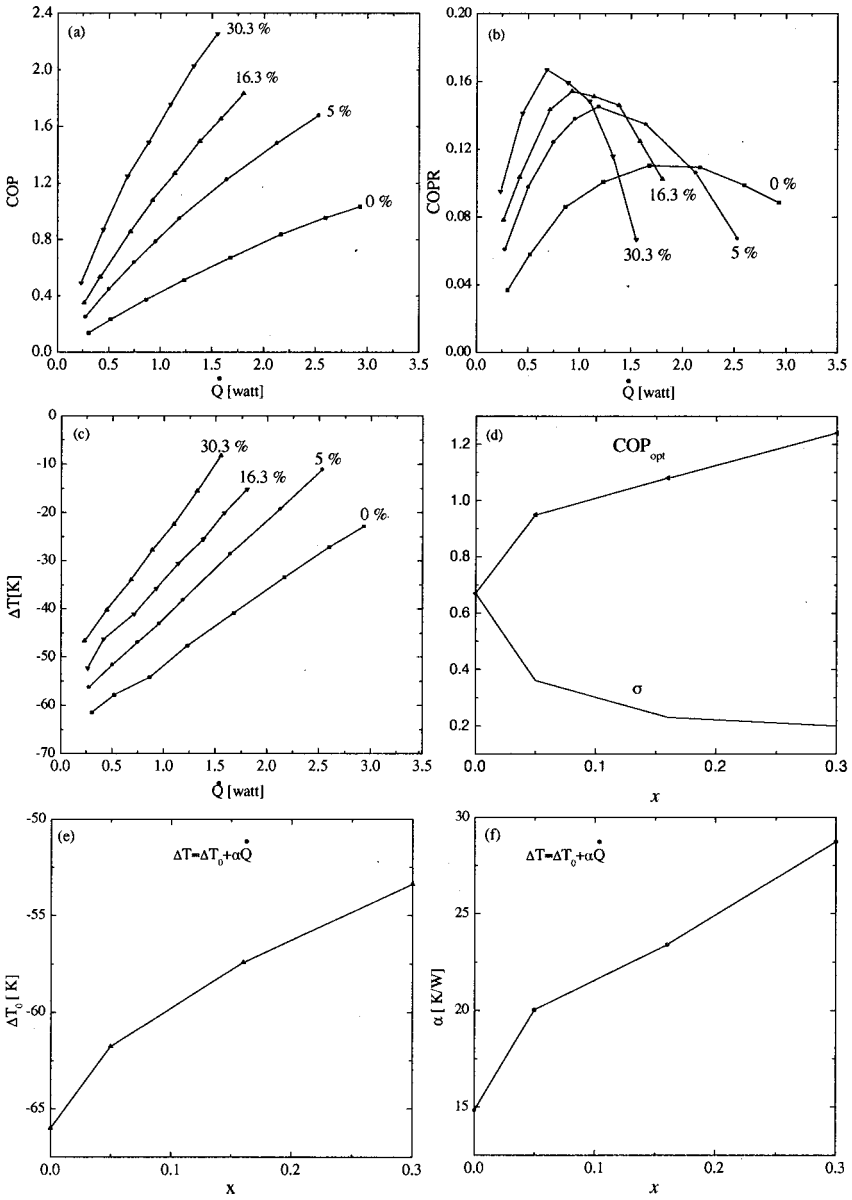


FIG. 8. Measurements with helium–xenon mixtures. The average pressure is 10 bar, the drive ratio is 1.4%, and a stack with a plate spacing of 0.3 mm is used. In (a), (b), and (c) the values of x are indicated at the corresponding curves. (a) COP. (b) COPR. (c) ΔT. (d) Optimal COP corresponding to the peak COPR as function of the mole fraction xenon x. (e) ΔT_0 as function of x. (f) α as function of x.

as functions of x. The slopes of ΔT_0 and α are steeper in the case of helium–xenon than in the case of helium–krypton. This indicates that the cooling power is more influenced by the addition of xenon than by the addition of krypton. This again is due to the density which is higher in the case of xenon.

A maximum COPR of 17% is reached with the mixture having the lowest Prandtl number and thus the minimal viscous losses in the system. This is equivalent to an improvement of 70% in comparison with pure helium. But, again the cooling power is smaller.

From the mixture measurements discussed above, we conclude that the Prandtl number can be decreased using helium and other noble-gas mixtures. This results in an improvement of the COPR of the refrigerator. However, at the same time the density increases which causes the cooling power to decrease. The behavior of the optimal performance $\mathrm{COP}_{\mathrm{opt}}$ as function of the noble gas mole fraction is similar to the one predicted by the calculations (Fig. 5).

In Table I, the linear fit results of the parameters ΔT_0 and α as a function of the mole fraction x for the different gas mixtures are given, according to the formula $\Delta T_0 = a + bx$ and $\alpha = c + dx$. The parameter a, which corresponds to ΔT_0 for pure helium, is the same for all mixtures, as it should be. The slope b increases as function of the molecular weight of the noble-gas component. This indicates the negative effect of the density on the cooling power. Likewise, the parameter c, which corresponds to the pure helium case for all mixtures, is the same for the three mixtures. The slope d increases as a function of the molecular weight of the noble component, which indicates again the negative effect of the density on the cooling power. In conclusion, we can say that

TABLE I. Linearly fitted data for the different gas mixtures.

Gas	a	b	c	d
Ar	-68 ± 1	19 ± 2	$c=16\pm0.5$	23 ± 1
Kr	-65 ± 1	41 ± 6	$c=15\pm1$	31 ± 6
Xe	-65 ± 1	40 ± 6	$c=16\pm1$	43 ± 7

TABLE II. Lennard-Jones potential parameters and collision integrals of noble gases.

Gas	ϵ/k (K)	ζ (Å)	$T=200$ K $\Omega^{(22)\star}$	$T=250$ K $\Omega^{(22)\star}$	$T=300$ K $\Omega^{(22)\star}$
He	10.2	2.6	0.75	0.72	0.70
Ne	34.9	2.8	0.90	0.87	0.84
Ar	122	3.4	1.26	1.16	1.09
Kr	171	3.6	1.47	1.32	1.23
Xe	221	4.1	1.67	1.49	1.37

the best COPR is obtained with the heaviest noble-gas component, but the cooling power is the smallest.

VII. CONCLUSIONS

The influence of the Prandtl number on the performance of the refrigerator is studied systematically by using mixtures of helium and other, heavier noble gases. The theoretical and experimental results show that the coefficient of performance improves as the Prandtl number decreases. However, the cooling power decreases when the mole fraction of the heavy noble-gas component increases in the mixture. It is shown that this decrease is a consequence of the increase of the density. A maximum performance relative to Carnot of 17% is obtained by using a mixture of helium and xenon containing 30% xenon. This is an improvement of 70% in comparison with pure helium. This trade-off between the performance and the cooling power has to be considered in the design of a thermoacoustic refrigerator.

APPENDIX A: KINETIC THEORY

In order to calculate the Prandtl number, Eq. (1), for gas mixtures we need the kinetic expressions for the dynamic viscosity μ, and thermal conductivity K. We now summarize the formulas for μ, K to be used with the tabulations of $\Omega^{(1,s)}$ for nonpolar molecules. These are written in the forms most convenient for practical calculations. In the text of Hirschfelder *et al.*,[7] it is indicated that the viscosity can be calculated to within 2% to 5% over a range of 400 K. But, the thermal conductivity may be in error as much as 10%, as this property is highly sensitive to the form of the potential function which is used.

1. Viscosity

The viscosity of a pure gas is given in Ref. 7, Eq. (8.2-18), as

$$\mu = 2.67\cdot 10^{-6}\frac{\sqrt{MT}}{\zeta^2\Omega^{(2,2)\star}(T^*)}, \tag{A1}$$

where M is the molecular weight in kg/kmol, T is the temperature in K, ζ is the collision diameter in Å, and $\Omega^{(2,2)\star}(T^*)$ is a tabulated function of the dimensionless temperature $T^*=kT/\epsilon$. These functions $\Omega^{(2,2)\star}(T^*)$ are given in Table II for noble gases at different temperatures. The unit of the viscosity is [Pas].

In order to discuss the viscosity of a binary gas mixtures, consisting of two components 1 and 2, it is convenient to define a quantity μ_{12} as

$$\mu_{12} = 2.67\cdot 10^{-6}\frac{\sqrt{\frac{2M_1M_2}{M_1+M_2}T}}{\zeta_{12}^2\Omega_{12}^{(2,2)\star}(T_{12}^*)}. \tag{A2}$$

This quantity can be regarded as the viscosity of a hypothetical pure substance, the molecules of which have molecular weight $2M_1M_2/(M_1+M_2)$, and interact according to the potential specified by the interaction parameters ζ_{12} and ϵ_{12}. The interaction of two nonpolar molecules of species 1 and 2 is described by the Lennard-Jones potential, where the parameters ζ and ϵ are given by the combining laws [Ref. 7, Eq. (3.6-8)]

$$\zeta_{12} = \tfrac{1}{2}(\zeta_1+\zeta_2), \tag{A3}$$

$$\epsilon_{12} = \sqrt{\epsilon_1\epsilon_2}, \tag{A4}$$

and are given in Table III. The appropriate parameters ζ_i and ϵ_i, for the molecules of the ith components, are given in Table II.

The viscosity μ_{mix} of a binary gas mixture, in terms of the quantities defined previously, is given by [Ref. 7, Eq. (8.2-22)]

$$\mu_{\mathrm{mix}} = \frac{1+Z_\mu}{X_\mu+Y_\mu}, \tag{A5}$$

where X_μ, Y_μ, and Z_μ are given by

$$X_\mu = \frac{x_1^2}{\mu_1}+\frac{2x_1x_2}{\mu_{12}}+\frac{x_2^2}{\mu_2},$$

TABLE III. Lennard-Jones potential parameters and collision integrals for gas mixtures.

Gas	ϵ_{12}/k (K)	ζ_{12} (Å)	$T=200$ K $\Omega^{(22)\star}$	$T=250$ K $\Omega^{(22)\star}$	$T=300$ K $\Omega^{(22)\star}$
He–Ne	18.89	2.67	0.82	0.79	0.77
He–Ar	35.31	2.98	0.91	0.87	0.84
He–Kr	41.80	3.08	0.93	0.89	0.87
He–Xe	47.52	3.33	0.96	0.92	0.89

TABLE IV. $A^\star$ and $B^\star$ parameters for binary gas mixtures at different temperatures.

	$T=200$ K		$T=250$ K		$T=300$ K	
Gas	$A^\star_{12}$	$B^\star_{12}$	$A^\star_{12}$	$B^\star_{12}$	$A^\star_{12}$	$B^\star_{12}$
He–Ne	1.1106	1.0950	1.1128	1.0944	1.1153	1.0909
He–Ar	1.1024	1.0909	1.1054	1.0916	1.1081	1.0912
He–Kr	1.1001	1.0941	1.1032	1.0903	1.1056	1.0911
He–Xe	1.0956	1.0978	1.1011	1.0916	1.1039	1.0918

$$Y_\mu=\frac{3}{5}A^\star_{12}\left[\frac{x_1^2}{\mu_1}\frac{M_1}{M_2}+\frac{2x_1x_2}{\mu_{12}}\frac{(M_1+M_2)^2}{4M_1M_2}\frac{\mu_{12}^2}{\mu_1\mu_2}+\frac{x_2^2}{\mu_2}\frac{M_2}{M_1}\right] \tag{A6}$$

$$Z_\mu=\frac{3}{5}A^\star_{12}\left[x_1^2\frac{M_1}{M_2}+2x_1x_2\left[\frac{(M_1+M_2)^2}{4M_1M_2}\left(\frac{\mu_{12}}{\mu_1}+\frac{\mu_{12}}{\mu_2}\right)-1\right]+x_2^2\frac{M_2}{M_1}\right].$$

The parameter $A^\star_{12}$ is a function of kT/ϵ_{12}, and defined as follows:

$$A^\star_{12}=\frac{\Omega^{(2,2)\star}}{\Omega^{(1,1)\star}}. \tag{A7}$$

These parameters are given for helium–other noble gases mixtures in Table IV.

2. Thermal conductivity

The thermal conductivity for a pure gas is given by [Ref. 7, Eq. (8.2–31)]

$$K=8.32\cdot10^{-2}\frac{\sqrt{T/M}}{\zeta^2\Omega^{(2,2)\star}(T^*)}=\frac{15}{4}\frac{R}{M}\mu, \tag{A8}$$

where R is the universal gas constant. The other parameters are as defined for viscosity. The unit is [W/Km].

In order to derive the thermal conductivity of binary gas mixture, it is convenient to define a quantity K_{12} as follows:

$$K_{12}=8.32\cdot10^{-2}\frac{\sqrt{\frac{M_1+M_2}{2M_1M_2}T}}{\zeta_{12}^2\Omega_{12}^{(2,2)\star}(T^*_{12})}. \tag{A9}$$

The thermal conductivity $K_{\rm mix}$ of a binary gas mixture is given by [Ref. 7, Eq. (8.2–36)]

$$K_{\rm mix}=\frac{1+Z_{\rm K}}{X_{\rm K}+Y_{\rm K}}, \tag{A10}$$

where $X_{\rm K}$, $Y_{\rm K}$, and $Z_{\rm K}$ are given by

$$X_{\rm K}=\frac{x_1^2}{K_1}+\frac{2x_1x_2}{K_{12}}+\frac{x_2^2}{K_2}$$

$$Y_{\rm K}=\frac{x_1^2}{K_1}U^{(1)}+\frac{2x_1x_2}{K_{12}}U^{(\rm Y)}+\frac{x_2^2}{K_2}U^{(2)}$$

$$Z_{\rm K}=x_1^2U^{(1)}+2x_1x_2U^{(\rm Z)}+x_2^2U^{(2)}$$

$$U^{(1)}=\frac{4}{15}A^\star_{12}-\frac{1}{12}\left(\frac{12}{5}B^\star_{12}+1\right)\frac{M_1}{M_2}+\frac{1}{2}\frac{(M_1-M_2)^2}{M_1M_2} \tag{A11}$$

$$U^{(2)}=\frac{4}{15}A^\star_{12}-\frac{1}{12}\left(\frac{12}{5}B^\star_{12}+1\right)\frac{M_2}{M_1}+\frac{1}{2}\frac{(M_2-M_1)^2}{M_1M_2}$$

$$U^{(\rm Y)}=\frac{4}{15}A^\star_{12}\frac{(M_1+M_2)^2}{4M_1M_2}\frac{K_{12}^2}{K_1K_2}-\frac{1}{12}\left(\tfrac{12}{5}B^\star_{12}+1\right)-\frac{5}{32A^\star_{12}}\left(\frac{12}{5}B^\star_{12}-5\right)\frac{(M_1-M_2)^2}{M_1M_2}$$

$$U^{(\rm Z)}=\frac{4}{15}A^\star_{12}\left[\frac{(M_1+M_2)^2}{4M_1M_2}\left(\frac{K_{12}}{K_1}+\frac{K_{12}}{K_2}\right)-1\right]-\frac{1}{12}\left(\frac{12}{5}B^\star_{12}+1\right),$$

where the parameter $B^\star_{12}$ is a function of kT/ϵ_{12}, and it is defined in terms of the $\Omega^{(1,s)\star}$ integrals as follows [Ref. 7, Eqs. (8.2–15)–(8.2–16)]:

$$B^\star_{12}=\frac{5\Omega^{(1,2)\star}-4\Omega^{(1,3)\star}}{\Omega^{(1,1)\star}}. \tag{A12}$$

These parameters are given for helium–noble gases mixtures in Table IV.

3. Specific heat and sound velocity

The last parameter which is needed to calculate the Prandtl number, using Eq. (1), for binary gas mixtures is c_p. Considering the binary gas mixture as an ideal gas, the specific heat at constant volume and constant pressure are given, respectively, by

$$c_v=\frac{\frac{3}{2}R}{M},\quad c_p=\frac{\frac{5}{2}R}{M}, \tag{A13}$$

where M is the apparent molecular mass of the mixture given by

$$M=x_1M_1+x_2M_2. \tag{A14}$$

The ratio of the isobaric to isochoric specific heats for the inert gas mixture is given by

$$\gamma=\frac{c_p}{c_v}=\frac{5}{3}. \tag{A15}$$

The sound velocity a in the binary gas mixture is given by

$$a^2=\gamma\frac{p}{\rho}=\gamma\frac{RT}{x_1M_1+x_2M_2}, \tag{A16}$$

where p is the pressure, ρ is the density, and T is the temperature of the gas mixture.

By substituting Eqs. (A5), (A10), and (A13) into Eq. (1), the Prandtl number of the binary gas mixtures can be calculated. The resultant expression is too long to be given explicitly here. This was incorporated in a computer program to accomplish the calculations.

[1] N. Rott, ''Thermoacoustics,'' Adv. Appl. Mech. **20**, 135–141 (1980).
[2] G. W. Swift, ''Thermoacoustic engines,'' J. Acoust. Soc. Am. **84**, 1146–1180 (1988).
[3] M. E. H. Tijani, ''Loudspeaker-driven thermo-acoustic refrigeration,'' Ph.D. dissertation, unpublished, Eindhoven University of Technology, 2001.
[4] F. W. Giacobbe, ''Estimation of Prandtl numbers in binary mixtures of helium and other noble gases,'' J. Acoust. Soc. Am. **96**, 3568–3580 (1994).
[5] J. R. Belcher, W. V. Slaton, R. Raspet, H. E. Bass, and J. Lightfoot, ''Working gases in thermoacoustic engines,'' J. Acoust. Soc. Am. **105**, 2677–2684 (1999).
[6] M. E. H. Tijani, J. C. H. Zeegers, and A. T. A. M. de Waele, ''Construction and performance of a thermoacoustic refrigerator,'' Cryogenics **42**, 59–66 (2002).
[7] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, New York, 1954), Chapters 1, 7, and 8.
[8] Reference 2, Eqs. (76) and (80).

Parametric dependencies for photoacoustic leak localization

Serdar H. Yönak[a] and David R. Dowling[b]
Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109-2121

(Received 14 September 2001; revised 29 March 2002; accepted 16 April 2002)

Unintended gas or liquid leaks from manufactured components or manufacturing systems may be detrimental to consumers, manufacturers, and the environment. Thus, leak testing is important for quality, safety, and environmental reasons. This paper describes parametric dependencies for photoacoustic leak localization. The technique is based on the interaction of 10.6-micrometer radiation from a carbon dioxide (CO_2) laser and a photoactive tracer gas, sulfur hexafluoride (SF_6). For the current investigations, acoustic signals are generated by scanning a laser beam at high speed through gas plumes formed above calibrated leaks. These signals are remotely measured with a four-microphone linear array and analyzed using Bartlett and minimum-variance-distortionless (MVD) matched-field processing (MFP) techniques to determine leak location. This paper extends prior work in photoacoustic leak testing through (i) use of more signal frequencies; (ii) parametric study of four different laser scan rates; and (iii) examination of mismatch between the actual acoustic environment and the propagation model used in the MFP; and (iv) presentation of leak localization results on a curved surface. For a 12-watt CO_2 laser exciting the small SF_6 gas plume produced by a one-cm^3-per-day leak with microphones placed 0.41 m from the leak location, root-mean-square localization uncertainties as small as ± 0.5 mm on a line scan of 0.46 m can be achieved when the largest possible number of signal frequencies fall in a measurement bandwidth of approximately 70 kHz. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1487838]

PACS numbers: 43.38.Zp, 43.35.Ud, 43.60.Gk [SLE]

I. INTRODUCTION

Many manufactured components or manufacturing systems have seals, internal chambers, or passages that contain or separate special liquids or gases from either the external environment or other portions of the component or system. Leaks that allow unintended flow of the special liquids or gases to beyond their expected domains often represent a serious quality control issue. Unintended leaks may degrade not only product performance and lifetime, they may also present safety concerns when the special gas or liquid is combustible, poisonous, or harmful to the environment. Unfortunately, common leak-testing techniques are either unable to determine leak location (mass spectrometer systems), are not sufficiently sensitive for many modern applications (dunk tanks, backscatter absorption systems, pressure decay methods), or are time and material intensive (ammonia–phenolphthalein suction systems) or are too reliant on operator skill (sniffing devices). Moreover, these techniques are not readily automated nor are they well suited to a factory environment where dust, smoke, oils, and environmental variations may lead to fouling and performance degradation. This paper presents new performance results for the emerging technique of photoacoustic leak testing which may overcome many of the limitations of prior systems, and has been shown to provide remote, noncontacting leak detection and localization in a matter of seconds (Yönak and Dowling, 1999).

The phenomenon of photoacoustics is the production of sound from amplitude modulated light. It was discovered more than a century ago (Bell, 1880, 1881) and has been an important research tool in spectroscopy and trace gas detection (see Kreuzer, 1977; Rosencwaig, 1980; Kreuzer and Patel, 1971; Claspy, 1977; Dewey, 1974; Dewey *et al.*, 1973; Goldan and Goto, 1974). More recently, Brassington (1982); McRae and Dewey (1992); McRae (1994); and Yönak and Dowling (1999, 2001) have exploited the photoacoustic effect for remote detection of gases and leaks.

The photoacoustic effect can be described in terms of energy transfer. A photon will be absorbed by a molecule when its wavelength matches a resonance peak in the molecule's absorption spectra. This process allows radiant energy to be transferred to the molecule, elevating its energy state (typically a change in vibrational or rotational quantum state). If the energized molecule is present in air at room temperature and pressure, molecular collisions provide abundant opportunities for it to transfer its excess energy to other molecules and thereby convert its excess energy into random translational motion, e.g., heat. This molecule-to-molecule energy transfer occurs on molecular time scales and can be regarded as essentially instantaneous for leak-testing purposes. Thus, the photoactive gas allows a remote light source to heat a distant gas cloud. When the intensity of the radiant light source is modulated, the rate of heat transfer to the absorbing molecules is also modulated and their temperature fluctuates. Because of the perfect gas law, gas-phase temperature fluctuations are converted to pressure and density fluctuations that may propagate away from the region of light absorption as sound waves. A general mathematical description of this process is available in Morse and Ingard (1968). A more specialized treatment emphasizing photoacoustic leak testing is provided in Yönak (2000).

[a]Currently employed by Ford Motor Company.
[b]Electronic mail: drd@engin.umich.edu

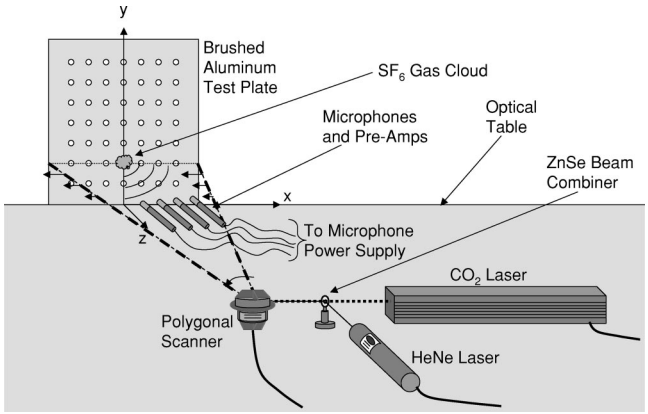


FIG. 1. Schematic of the experimental setup. The CO_2 laser is combined with a He–Ne laser and then reflected from a rotating polygonal mirror. The combined laser beams reflect from the scanner and traverse a flat plate on which calibrated leaks can be mounted. Photoacoustic sound produced by the leaking tracer gas is recorded by a four-element linear broadside microphone array. These recorded photoacoustic signals are processed with matched-field techniques to determine leak location.

The photoacoustic effect can be used to determine the location of leaks by precharging the test component with a photoactive tracer gas, scanning the component with a beam of laser light whose wavelength is matched to the absorption band(s) of the tracer gas, and recording any photoacoustic sounds. If a leak is present, the tracer gas will form a plume near the leak location and the plume will emit photoacoustic sound when it is illuminated by the laser beam. If this photoacoustic sound is recorded with one or more microphones in a known environment, matched-field processing (see Appendix A) may be used to determine the likely leak location(s). Such a leak localization scheme has been previously reported by Yönak and Dowling (1999). This paper extends their results to (i) use of more signal frequencies; (ii) parametric study of four different laser scan rates; (iii) examination of mismatch between the actual environment and the environmental model used in the MFP; and (iv) leak localization on a curved surface. The new findings are important because they suggest that detailed knowledge of the acoustic environment is not necessary for acceptable-accuracy leak localization as long as all the microphones receive direct-path photoacoustic sound from the laser-illuminated tracer gas plume above the leak.

The remainder of the paper is divided into three sections and two appendices. Section II briefly describes the experimental setup. Section III presents the main experimental results. The final section provides a summary of this paper and the conclusions drawn from this research. Appendix A briefly covers the mathematical description of the two MFP techniques utilized for this paper and provides several references. Appendix B describes the field models employed in these studies.

II. EXPERIMENT

The experimental setup is largely the same as that described in Yönak and Dowling (1999), so the following summary is fairly terse. Full experimental details are provided in Yönak (2000).

The current experimental setup, shown in Fig. 1, is for one-dimensional detection and localization of leaks along a line parallel to the x axis on a flat aluminum plate mounted on top of and perpendicular to an optical table. For some experiments the plate was replaced by a vertical cylinder with its axis parallel to the y axis. The x- and z-coordinates of this leak are unknowns in the signal-processing routines. The laboratory environment, consisting primarily of painted cinder-block walls with a concrete ceiling and a tile floor, was not acoustically treated. Laboratory air temperature in the vicinity of the experiments was monitored to $\pm 1\,^{\circ}\mathrm{C}$ using a bare-wire thermocouple.

For these experiments, the laser beam from a Synrad grating-tuned CO_2 laser delivering 12 watts at 10.6 micrometer was scanned through a plane 10.5 cm above and parallel to the optical table top. Two calibrated leaks (from Vacuum Instrument Corporation) of the tracer gas, SF_6, having volume flow rates of 1.19×10^{-5} cm^3/s (approximately 1 cm^3 per day) and 1.66×10^{-6} cm^3/s (approximately 1-cm^3 per week) were mounted in threaded holes in the test plate at the height of the laser beam scan. The 10.6-μm CO_2-laser radiation matches a strong absorption line of the SF_6 tracer gas (Lymann *et al.*, 1986), making this laser–gas pair well suited for photoacoustic sound production. The nominal CO_2-laser beam diameter ($1/e^2$) at the plate surface was 8.0 ± 0.5 mm. Since the CO_2 laser is invisible, it was combined with a 5-mW red He–Ne laser using a Zn–Se beam combiner to visualize the potentially hazardous CO_2 beam. The combined beams were incident on a Lincoln Laser beam scanner consisting of a 12 sided polygonal mirror mounted on the shaft of a high-speed motor capable of 37 500 RPM. The scanner creates a line scan centered at normal incidence on the aluminum test plate or test cylinder located 0.84 m away. The nominal laser beam scan rate for these experiments was 6.25 kHz; however, results for several different scan rates are presented in the next section.

When the laser beam illuminates the leaked tracer-gas plume, photoacoustic sound is generated at frequencies that are harmonics of the scan rate. This photoacoustic sound was measured using four Brüel & Kjær Model 4135 $\frac{1}{4}$-in. broadband (up to $\sim$80 kHz) microphones spaced 25.4 ± 0.5 mm apart in a broadside horizontal linear array set 0.41 m from the test plate and 14.6 cm above the table top. For the experiments with the vertical cylinder, the test geometry was the same, except the microphone array was placed 20.5 cm away from the cylinder axis. The signals from these microphones were amplified and bandpass filtered between 3 and 110 kHz before being 12-bit digitized at 312.5 kHz via a Data Translation DT 3010 A/D board mounted in a PC-type computer. At this data rate, 65 536 samples per channel were collected in approximately 0.21 s for a complete experiment. The digitized photoacoustic signals were Fourier analyzed, and comb filtered to extract the complex Fourier amplitudes $M_i(\omega_n)$ (i is an index for microphone number) at the signal frequencies ω_n (in rad/s, n is an index for signal harmonic number). These amplitudes and field models of the acoustic environment were employed for multiple-frequency linear Bartlett MFP and nonlinear or adaptive minimum-variance distortionless (MVD) MFP to determine the leak location.

The mathematical formulation of the MFP routines is given in Appendix A along with references. The formulation of the requisite field models is given in Appendix B. The output of the MFP routines, $B_B(\mathbf{r})$ and $B_{\mathrm{MVD}}(\mathbf{r})$ for the Bartlett and MVD processors, respectively, are ambiguity surfaces that peak at the likely location(s) of the unknown sound source(s). Here, ambiguity surface peaks also indicate the leak location because molecular diffusion ensures that the tracer gas plume (the photoacoustic sound source) will be concentrated near the leak location (Yönak and Dowling, 1999).

III. EXPERIMENTAL RESULTS

This section presents the main leak localization findings in four subsections. The first subsection establishes the baseline localization performance of the system with a scan rate of 6.25 kHz when 12 signal frequencies are used. The second shows the results of a parametric study of how the scan rate influences the localization accuracy and the peak-to-sidelobe ratio. The third subsection addresses mismatch between the actual and model acoustic environments. The final subsection provides localization results for leaks placed on a cylindrical surface.

A. Basic localization

The intersection of the laser beam scan with the flat plate shown schematically in Fig. 1 is a line parallel to the x axis. Figure 2 shows the 12-frequency (6.25 to 75.0 kHz) MFP ambiguity surfaces from the two processors along this line when the 1.19×10^{-5} cm^3/s and 1.66×10^{-6} cm^3/s leaks are present separately at the center of the plate ($x=0$) and when the leak is absent. Here, the rectangular-corner field model described in Appendix B is used in the MFP routines. Both processors show a clear peak at the correct location when the leaks are present. As expected, the MVD processor peak is narrower and the sidelobes are lower. These results are very similar but somewhat better than the eight-frequency results in Yönak and Dowling (1999).

Interestingly, both processors provide weaker false peaks at $x=0$ when the leak is absent. These false peaks were the subject of further investigation because they determine the sensitivity limit below which noise or other unintended photoacoustic sources will appear as a low-flow-rate leak in the MFP outputs. An examination of the experiment determined two possible origins for the false MFP peaks: noise from the spinning polygonal mirror, and unintended photoacoustic sound produced by the interaction of the CO_2 laser beam with the surface of the brushed aluminum test plate. The spinning polygonal mirror's facets produce sound at the first few signal frequencies and the scanner itself is located directly behind the linear microphone array. This geometrical configuration of array and spinning mirror allows the facet-produced sound to appear as a sound source directly in front of the microphone array. In addition, photoacoustic sound produced directly from the laser beam interacting with the test plate will occur at the signal frequencies and will arrive nearly in-phase across the short microphone array. The chosen geometrical configuration of

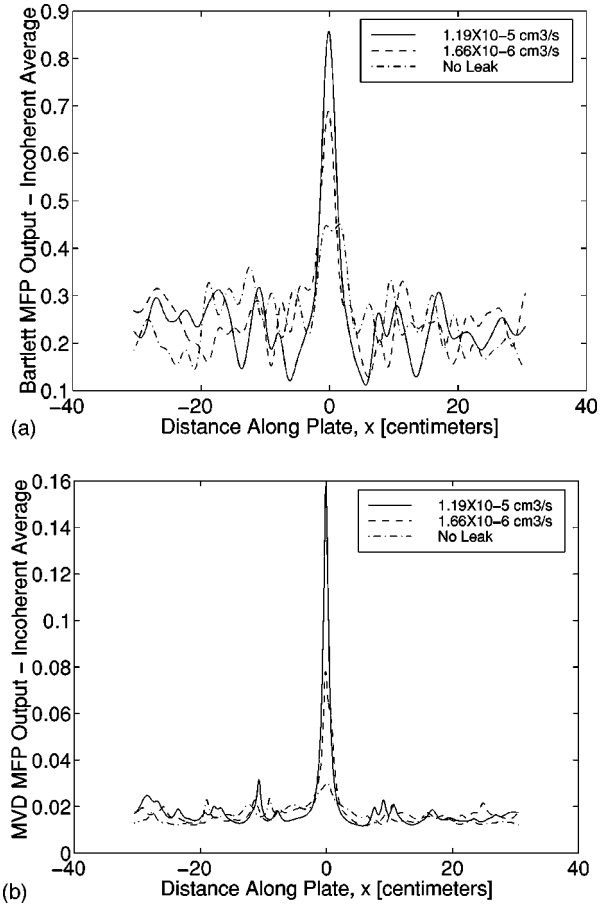


FIG. 2. Matched-field processing output for the Bartlett (a) and MVD (b) MFP processors along the line of contact between the laser beam and the test plate for the two calibrated leaks of this study separate placed at the center of the plate, and when the entire experiment is running but no leak is present on the plate. Laser scan rate = 6.25 kHz, and the first 12 signal harmonics were used in the MFP.

the array (each microphone equidistant from the test plate) causes the signal processing to interpret this unintended test-plate-generated photoacoustic sound as a sound source located near $x=0$.

To determine the importance of these potential unintended sound sources, they were eliminated from the experiment one at a time. Figure 3 shows both Bartlett and MVD output when the scanner and the laser are both on but: the leak is absent (solid lines), the polygonal mirror is spinning but the laser shutter is closed (dashed lines), and when the polygonal mirror is not spinning and the laser shutter is closed (dash-dot lines). The false peaks at $x=0$ in both processors are reduced to the sidelobe level away from $x=0$ when the mirror is not spinning and the laser is not touching the plate (the laser-shutter-closed condition). Thus, the false MFP peaks are caused by the two unintended sound sources described above. In addition, both processors provide a taller false peak when the laser shutter is closed with the scanner still spinning. This is consistent with some destructive interference occurring between the two unintended sound sources at the microphone array.

The 12-frequency localization results can also be displayed in the horizontal plane of the laser scan (the $x-z$ plane). Figures 4 and 5 show the Bartlett and MVD ambigu-

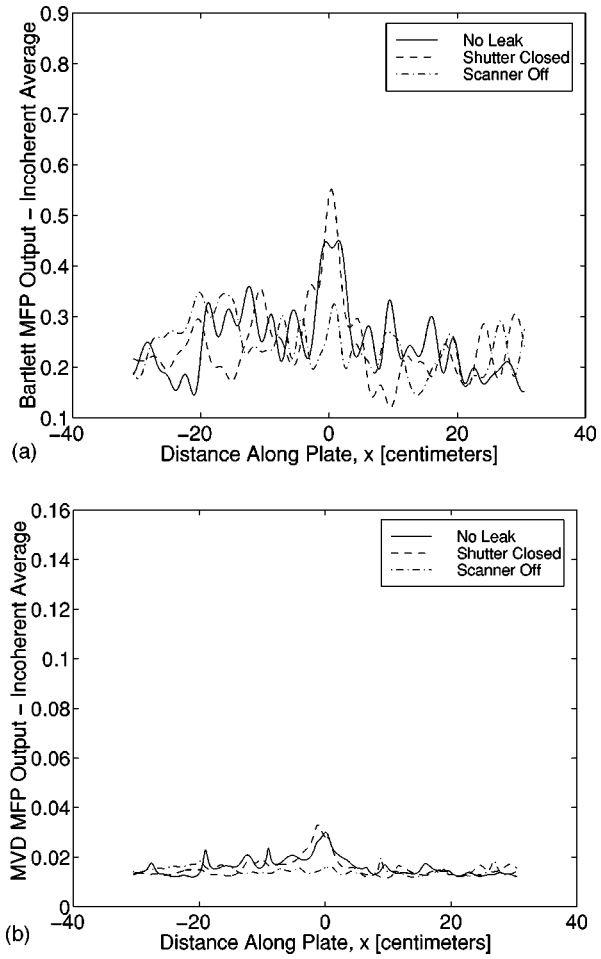


FIG. 3. Matched-field processing output for the Bartlett (a) and MVD (b) MFP processors along the line of contact between the laser beam and the test plate when the entire experiment is running but: no leak is present on the plate (———), no leak is present on the plate and the laser shutter is closed (---), and no leak is present on the plate, the laser shutter is closed, and rotating polygonal mirror is not spinning (-·-·-). Laser scan rate $=6.25$ kHz, and the first 12 signal harmonics were used in the MFP.

ity surfaces, respectively, when the 1.19×10^{-5} cm^3/s leak was placed at $x=-22.86$ cm (a); -15.24 cm (b), -7.62 cm (c); and 0 cm (d). Here, the rectangular-corner field model (Appendix B) is used in the MFP routines. The left-hand edge of each frame in each figure corresponds to the surface of the plate where it is illuminated by the laser beam. The processor outputs at the left edge of Fig. 4(d) and Fig. 5(d) are the same as the solid curves in Fig. 2. The small circles are centered on the actual location of the leak. In all cases the MFP peak is located at the leak location. When compared with the prior eight-frequency results obtained with a 20-facet polygonal mirror (see Yönak and Dowling, 1999, Figs. 8 and 9), the current results show better multidimensional localization away from $x=0$ and significantly reduced sidelobes.

B. Variable laser scan rate

An ideal photoacoustic leak localization system would indicate the location of a leak without error via a single narrow peak that is significantly higher than any sidelobe. The rate at which the laser beam scans across the test component is a parameter that can be varied to optimize localization performance. Such a parameter study was pursued using four scan rates covering an octave, and the 1.19×10^{-5} cm^3/s leak mounted at $x=0$ on the flat test plate. Ten trials for each scan rate were conducted to determine localization performance statistics for the Bartlett MFP processor. Here, root-mean-square (rms) localization error, processor peak full-width at half-maximum (FWHM), and ambiguity surface peak-to-highest sidelobe ratio were determined as a function of the data acquisition bandwidth.

The localization accuracy results are shown in Fig. 6 for scan rates of 3.75, 5.00, 6.25, and 7.50 kHz. In all cases, rms localization error in the x direction decreases with increasing bandwidth up to a frequency of 70 kHz or so, with the lower two scan rates tending to produce smaller rms errors. The most likely reason that larger bandwidths do not continue to improve localization accuracy is that the metrology available for these experiments could only reduce microphone placement uncertainty to about ±0.5 mm. At acoustic frequencies above 70 kHz (wavelengths less than 5 mm), these small microphone placement errors cause field-model phase errors which degrade the high-frequency MFP localization results.

The somewhat better performance of the smaller scan rates has a simple explanation. Because the signal frequencies are integer multiples of the scan rate, lower scan rates pack more signal harmonics into a fixed bandwidth. A scan rate of 3.75 kHz packs twice as many signal frequencies as a 7.5-kHz scan rate into any fixed bandwidth. Thus, when signal-to-noise ratios are good, localization accuracy is governed by the precision of the geometrical match between the field model at a given frequency and the number of signal harmonics that lie within the data acquisition bandwidth (more is better).

Figure 7 shows Bartlett MFP peak full-width at half-maximum (FWHM) results for the four scan rates as a function of data acquisition bandwidth for the 1.19×10^{-5} cm^3/s leak mounted at $x=0$ on the flat test plate. For all scan rates the curves are monotonically descending, but flatten out at about 4 cm for bandwidths greater than approximately 50 kHz. Because of the incoherent averaging of the narrow-band MFP results, the FWHM results in Fig. 7 do not necessarily scale with the wavelength of the highest-frequency signal component. In fact, an ideal computer simulation of the experiment produces essentially the same results as those shown in Fig. 7 except that the limiting high-bandwidth value of the Bartlett MFP peak FWHM is closer to 3 cm (the presumed diffraction limit). Thus, the localization resolution is set by the experimental geometry and imperfections (noise, microphone placement errors, etc.), and diffraction.

Figure 8 shows measured peak-to-highest sidelobe ratios (PTSLR) for the Bartlett processor for the four scan rates as a function of data acquisition bandwidth for the 1.19×10^{-5}-cm^3/s leak mounted at $x=0$ on the flat test plate. Once the bandwidth of this signal is sufficient for signal components combined with the microphone array geometry to produce sidelobes that correspond to locations on the test plate, the PTSLR level hovers near 3 for increasing bandwidth. As for the FWHM results, an idealized computer

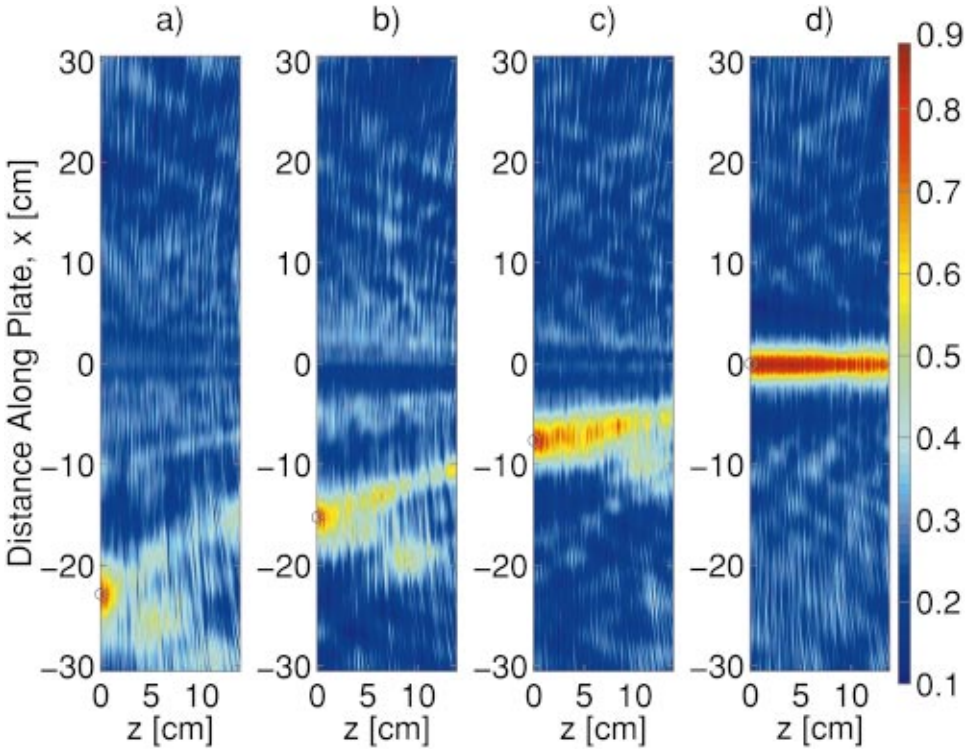


FIG. 4. Bartlett MFP ambiguity surfaces in the $x-z$ plane when the 1.19 $\times 10^{-5}$-cm^3/s leak was placed at $x = -22.86$ (a); -15.24 (b); -7.62 (c); and 0 cm (d). High ambiguity values indicate the likely location(s) of the leak. The plate lies at the left edge of each frame. The small circles are centered on the actual leak location. Laser scan rate = 6.25 kHz, and the first 12 signal harmonics were used in the MFP. The color bar indicates the magnitude of the Bartlett MFP output.

simulation of the experiment produces a plot that is very similar to Fig. 7 with a high-bandwidth PTSLR that hovers near 4. Thus, the PTSLR is also set by the experimental geometry and imperfections, and diffraction.

Overall, the findings for the MVD processor are either similar to those shown in Figs. 6–8 for the Bartlett processor or the MVD findings differ in ways that can be anticipated. The MVD processor's rms localization errors are comparable to the Bartlett values. The MVD peak width is generally smaller by a factor of 2 to 3. The MVD peak-to-sidelobe ratios are typically larger by a factor of approximately 1.5. These performance improvements are typical for the MVD processor when the signal-to-noise ratio (SNR) is relatively high. The SNR of the current measurements with the 1.19 $\times 10^{-5}$-cm^3/s leak is approximately 10–15 dB. However, the observed performance improvements for the MVD processor may be entirely lost at signal-to-noise ratios of unity or below (0 dB or less).

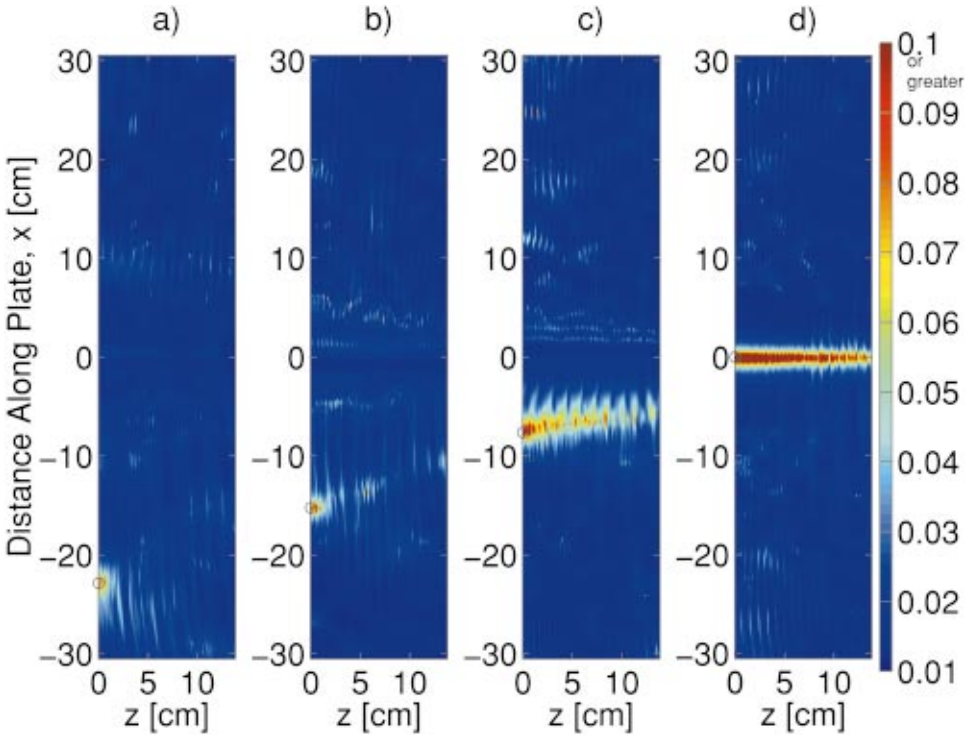


FIG. 5. Same as Fig. 4 except these results are from the MVD processor. The color bar indicates the magnitude of the MVD MFP output.

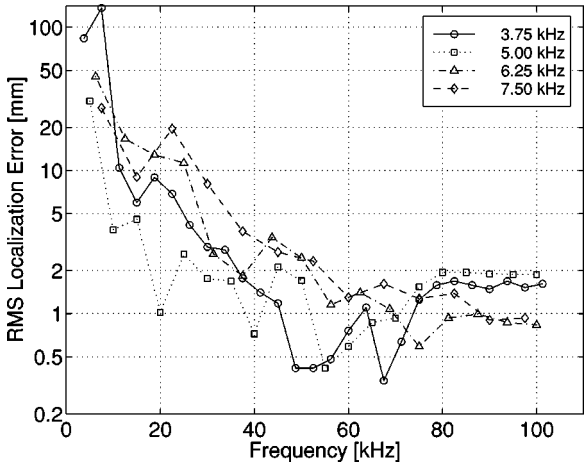


FIG. 6. Root-mean-square (rms) localization error for the Bartlett processor based on ten trials versus data acquisition bandwidth for four different laser scan rates when the 1.19×10^{-5}-cm^3/s leak was placed at the center of the plate $x=0$. Increasing signal bandwidth generally improves localization accuracy up to approximately 70 kHz.

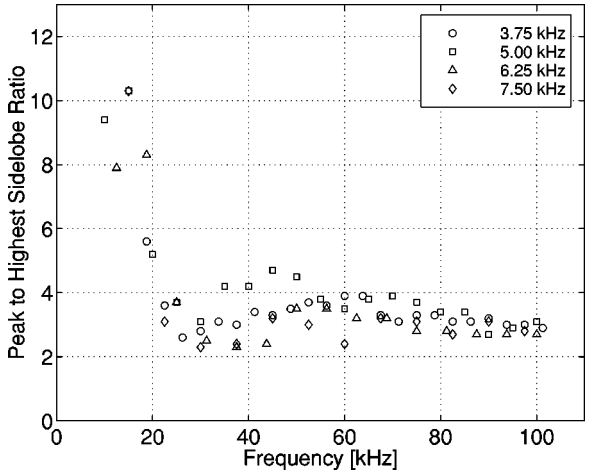


FIG. 8. Peak-to-highest sidelobe ratio (PTSLR) along the plate surface ($z=0$) versus data acquisition bandwidth for four different laser scan rates when the 1.19×10^{-5}-cm^3/s leak was placed at the center of the plate ($x=0$). The PTSLR hovers near 3 once the bandwidth is large enough for the microphone array to produce sidelobes that fall on the plate surface.

C. Sensitivity to environment-field model mismatch

To be truly applicable to a wide range of leak-testing situations involving heat exchangers, piping, valves, pressure vessels, vacuum chambers, gasket seals, etc., photoacoustic leak-testing systems will need to be easily configured for a variety of testing geometries. One way to approach this problem is to ignore acoustic reflections in the field model used in the MFP routines and base the array processing on a free-space propagation model (see Appendix B). The main advantage of this approach is that a free-space propagation model should be almost universally applicable to some part of the measured photoacoustic signal when the microphones have a clean line-of-sight path to the photoacoustic sound source. When a free-space field model is employed, MFP reduces to more ordinary near-field or spherical wave beamforming (see Ziomek, 1995).

The performance of this approach with known environment-model mismatch is documented here for photoacoustic measurements made with two reflecting surfaces (the table top and the test plate). Figures 9 and 10 are the same as Figs. 4 and 5 except that the field model (free-space) is mismatched to the actual acoustic environment (a rectangular corner). Comparing Figs. 9 and 10 to Figs. 4 and 5 shows that the localization results with the possibly universal free-space model might be acceptable even though the localization error is worse (as large as a cm or two) and the multidimensional localization is lost (both processors form beams for all leak locations). Comparisons between Figs. 9 and 10 show that the wider beams developed by the Bartlett processor (Fig. 9) intersect the plate surface ($z=0$) consistently closer to the actual leak location than the narrower beams for the MVD processor (Fig. 10). Thus, the Bartlett processor can be considered more robust to this type of environment-field model mismatch.

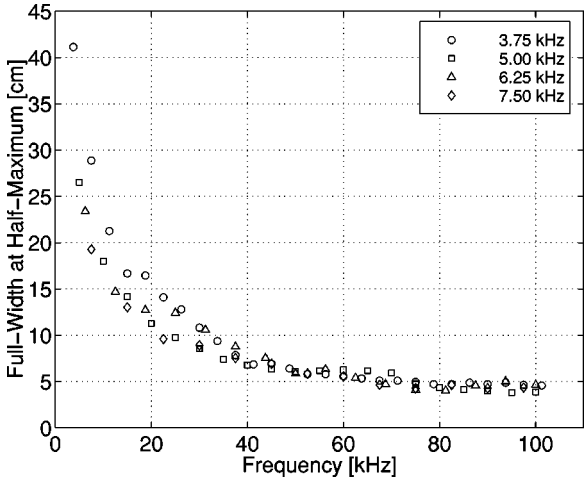


FIG. 7. Bartlett ambiguity peak full-width at half-maximum in the x direction versus data acquisition bandwidth for four different laser scan rates when the 1.19×10^{-5}-cm^3/s leak was placed at the center of the plate ($x=0$). Increasing signal bandwidth generally decreases the peak width (e.g., increases the system resolution) up to approximately 50 kHz.

D. Localization on curved surfaces

Up to this point, all of the reflecting surfaces in the actual acoustic environment were large and flat. In this subsection, photoacoustic leak localization on a cylindrical surface is considered. The geometry is similar to that shown in Fig. 1 except for a 5.1-cm radius cylinder whose axis is vertical replaces the flat test plate. Again, two field models were considered. The first is a geometrical acoustics (ray-theory) approximation to the full field model for a point source near a cylindrical reflecting surface (see Appendix B). The flat reflecting surface of the optical table top is retained in this ray-theory field model. The second is the free-space field model.

Figure 11 shows 12-frequency ambiguity surfaces for the Bartlett processor in the $x-z$ plane based on the approximate cylindrical geometry field model when the 1.19×10^{-5}-cm^3/s leak was placed on the cylindrical surface 0° (a); 30° (b); 60°(c), and 80° (d) from the line connecting the axis of the cylinder and the center of the microphone array. For the 0° and 30° cases, the leak is localization error is ±1

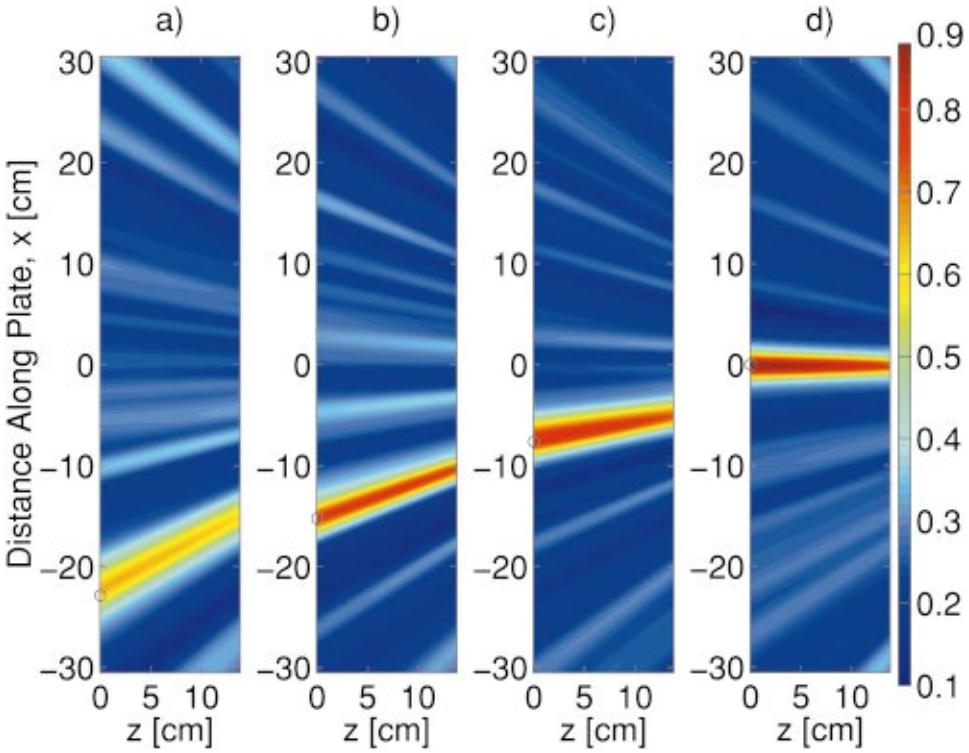


FIG. 9. Same as Fig. 4, except that the measurements are processed using the free-space field model.

mm. At 60°, the localization is off by a few mm but is still within the bound determined by the width of the localization peak. However, the leak is not localized for the 80° case. The localization failure in this final leak orientation is caused by acoustic shadowing of the microphones and an apparent change in the laser illumination of the leak. Together, these factors lead to a severe reduction in signal-to-noise ratio for the 80° case. The mismatch between the actual field the microphones measure and the approximate ray-theory-based field model, which does not account for shadowing, may also play a minor role in this loss of localization.

Figure 12 shows the same measurements as Fig. 11 except the Bartlett processing was done with the free-space field model. Here, the localization results for 0° [Fig. 12(a)]; 30° [Fig. 12(b)]; and 60° [Fig. 12(c)] are of essentially the same quality as those shown in Figs. 11(a)–(c). However, the 80° results [Fig. 12(d)] still fail to localize the leak.

Use of the MVD processor in this cylindrical geometry

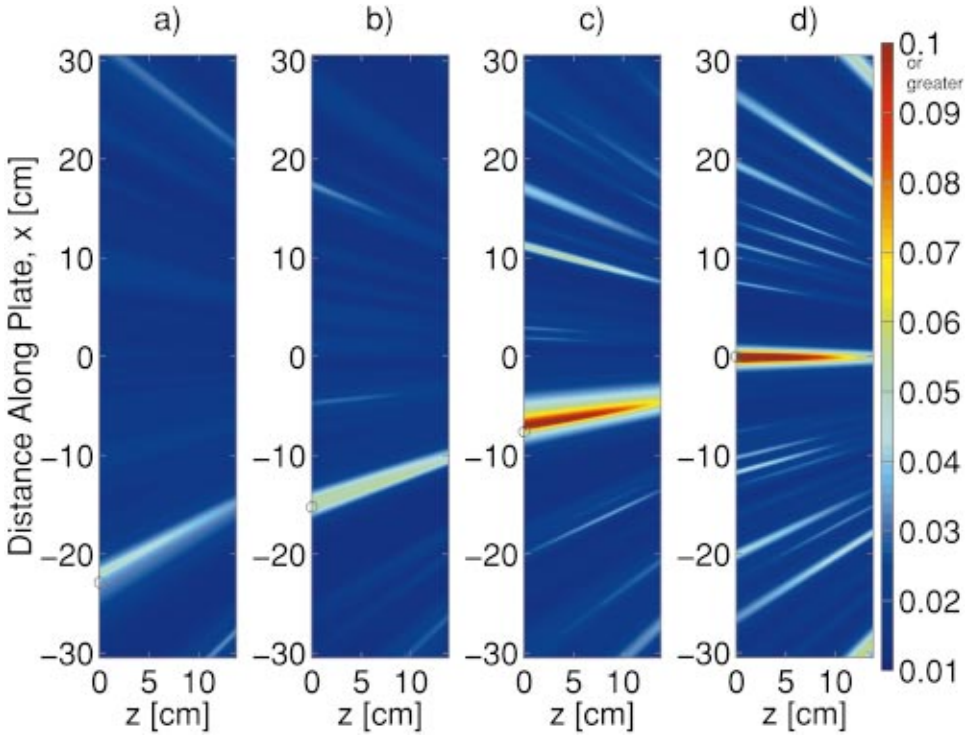


FIG. 10. Same as Fig. 5, except that the measurements are processed using the free-space field model.

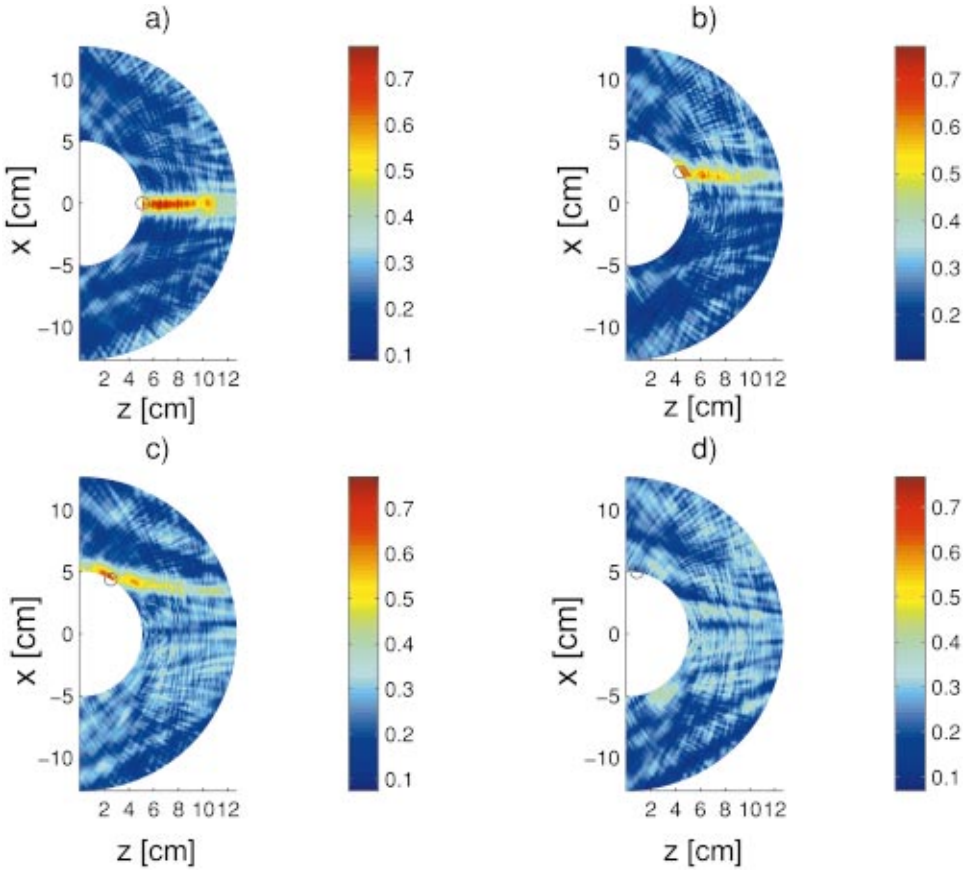


FIG. 11. Bartlett MFP ambiguity surfaces in the $x-z$ plane when the 1.19 $\times 10^{-5}$-cm^3/s leak was placed on a cylindrical surface at 0° (a); 30° (b); 60° (c); and 80° (d) from a line drawn between the center of the microphone array and the axis of the cylinder which is perpendicular to the $x-z$ plane. High ambiguity values indicate the likely location(s) of the leak. The cylindrical surface lies at the inner curved edge of each semiannulus. The small circles are centered on the actual leak location. The laser scan rate $=6.25$ kHz, and the first 12 signal harmonics were used in the MFP. The color bar indicates the magnitude of the Bartlett MFP output.

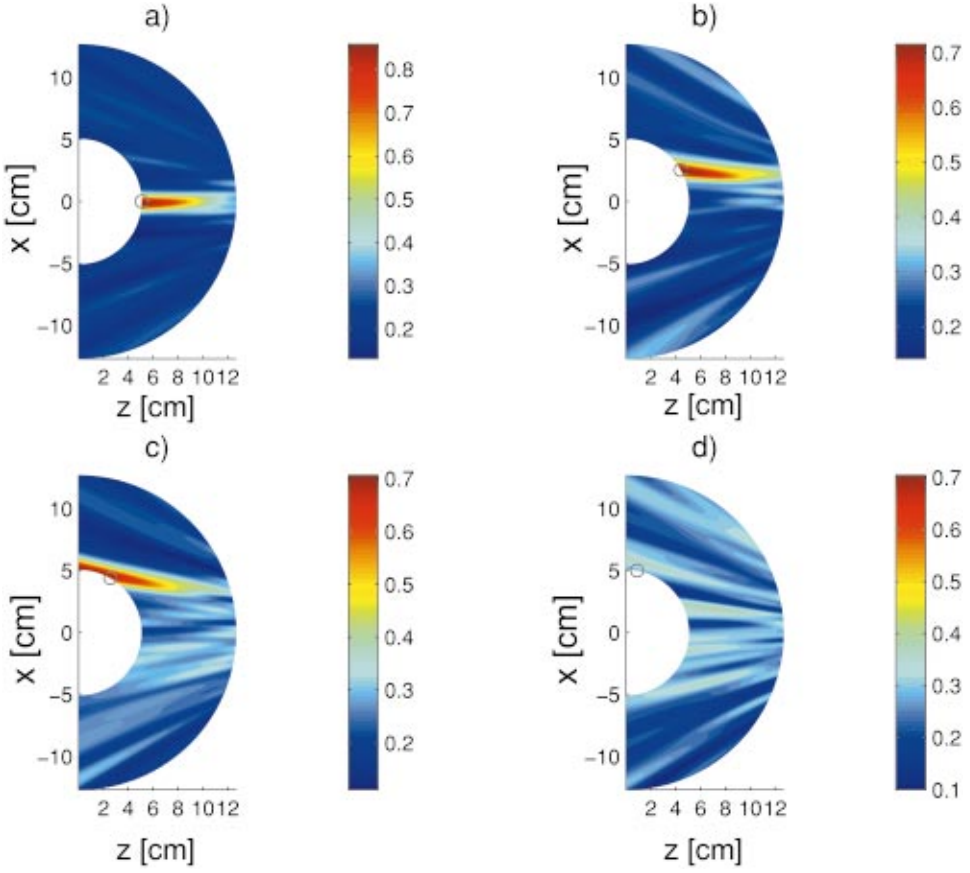


FIG. 12. Same as Fig. 11 except that the measurements are processed using the free-space field model.

produces the expected changes to the Bartlett processor results shown in Figs. 11 and 12: tighter localization and better sidelobe suppression when the signal-to-noise ratio is good. However, the MVD processor also fails to localize the 1.19 $\times 10^{-5}$-cm^3/s leak when it is placed 80° from the front of the cylinder.

IV. SUMMARY AND CONCLUSIONS

This paper has presented new results for the emerging technique of photoacoustic leak localization. Leak localization performance involving additional signal frequencies, different laser scan rates, mismatched field models, and curved surfaces have all been quantitatively investigated for leak rates of approximately 1 cubic centimeter per day. In addition, two sources of coherent noise that limit system sensitivity have been identified.

Three main conclusions can be drawn from this investigation. First, use of high-frequency signal components only improves the localization results when the field model employed in the matched-field processing is sufficiently accurate to properly predict signal-phase differences on the microphone array. For the current experimental setup, this localization accuracy limitation is reached at about 70 kHz. More refined metrology could easily push this frequency higher.

Second, lowering the laser scan rate to pack more signal frequencies into the available bandwidth generally improves localization performance when the signal-to-noise ratio is good. For a 70-kHz bandwidth, the 3.75-kHz scan rate measurements localized the 1-cubic-centimeter-per-day leak to within ± 0.5 mm, an uncertainty that is a factor of 2 or 3 smaller than that obtained at laser scan rates of 6.25 or 7.5 kHz. However, this performance trend cannot be followed indefinitely to lower scan rates because the photoacoustic signal amplitude also depends on the scan rate, and signal-to-noise problems will arise at scan rates that are too slow.

Third, based on the experiments involving leaks on both planar and curved surfaces, leak localization utilizing only direct-path sound propagation between the tracer gas plume and the microphones appears to be a viable way to extend photoacoustic leak testing to components having complex surfaces, provided that direct path sound is captured by the microphones. The results also show that use of an accurate field model improves localization results, but many practical applications of photoacoustic leak testing may involve complex shapes and a robust general-purpose field model may be best for these cases. Naturally, leaks on interior surfaces that are not illuminated by the scanning laser cannot be localized by photoacoustic means even if direct paths exist between the leak and the microphones. However, because the laser beam travels in a straight line, direct path photoacoustic sound from an illuminated leak to the microphones should be present as long as the microphones lie somewhere near the apparent origin of the scanning laser beam. Or, stated in active sonar terminology, the photoacoustic leak test system should be approximately monostatic (a predominantly backscatter-based system) as opposed to completely bistatic (a forward- or side-scatter-based system). Thus, if a leak can be properly illuminated, then photoacoustic means should be a robust way to localize it because direct path sound will occur along the direction of the laser illumination path.

And, as a final note, an interesting future possibility for this development effort is to attempt a coherent addition of the MFP results based on a deduced phase relationship between signal harmonics. Naturally, such an effort would require knowledge of the photoacoustic pulse shape, stability of this shape from pulse to pulse, and coherent MFP (Westwood, 1992).

ACKNOWLEDGMENTS

This research project was supported by Ford Motor Company, Advanced Manufacturing Technology Development. The advice of Dr. Thomas McRae is also appreciated.

APPENDIX A: MATCHED-FIELD PROCESSING

Matched-field processing (MFP) commonly refers to a closely related set of array signal-processing techniques for localization of remote unknown sound sources in known acoustic environments. The two MFP schemes employed in this paper, the Bartlett and minimum variance distortionless (MVD) processors, are described in Bucker (1976) and Baggeroer *et al.* (1988), respectively. Summaries and overviews of MFP are provided in Baggeroer *et al.* (1993), Tolstoy (1993), and Jensen *et al.* (1994).

For the present leak-testing application of MFP with m microphones located at $\mathbf{r}_i=(x_i,y_i,z_i)$, the narrow-band version of either the Bartlett processor or minimum variance distortionless (MVD) processor at a frequency ω (rad/s) and interrogation location $\mathbf{r}=(x,y,z)$, can be cast into the same form

$$B(\mathbf{r};\omega)=\sum_{i=1}^{m}\sum_{j=1}^{m} w^*(\mathbf{r},\mathbf{r}_i;\omega)K_{ij}(\omega)w(\mathbf{r},\mathbf{r}_j;\omega), \quad \text{(A1)}$$

where $B(\mathbf{r};\omega)$ is the processor output or ambiguity function, $w(\mathbf{r},\mathbf{r}_i;\omega)$ is the weighting for each microphone, $K_{ij}(\omega)$ is the measured cross-spectral density matrix of the microphone measurements $M_i(\omega)$ at frequency ω

$$K_{ij}(\omega)=\frac{M_i(\omega)M_j^*(\omega)}{\sum_{n=1}^{m}|M_n(\omega)|^2}, \quad \text{(A2)}$$

and the asterisk denotes complex conjugation. In some applications of MFP, $K_{ij}(\omega)$ is determined from an ensemble average of several independent sets of measurements (Baggeroer *et al.*, 1993).

For the Bartlett processor, the microphone weights are determined from

$$w_B(\mathbf{r},\mathbf{r}_i;\omega)=\frac{G(\mathbf{r},\mathbf{r}_i;\omega)}{\{\sum_{n=1}^{m}|G(\mathbf{r},\mathbf{r}_n;\omega)|^2\}^{1/2}}, \quad \text{(A3)}$$

where $G(\mathbf{r},\mathbf{r}_i;\omega)$ is the Green's function for the Helmholtz equation calculated for the known acoustic environment at frequency ω. Use of (A3) for the microphone weights in (A1) produces the Bartlett MFP output: $B_B(\mathbf{r};\omega)$. Bartlett MFP can be interpreted as the spatial correlation of the field at themicrophone array produced by a fictitious test source

located at $\mathbf{r}$ with the measured field from the unknown source. The normalizations shown in (A1)–(A3) confine $B_B(\mathbf{r};\omega)$ between zero and unity; $B_B(\mathbf{r};\omega)\ll 1$ implies little chance that the unknown source resides at $\mathbf{r}$ while $B_B(\mathbf{r};\omega)$ near unity implies a high probability the unknown source is located at $\mathbf{r}$. Bartlett MFP is considered linear because the measurements do not influence the weight vectors.

For the MVD processor the cross-spectral density of the measurements is used in the construction of the microphone weights to minimize the output of the processor for all locations except the location of interrogation, $\mathbf{r}$. Thus, the MVD weight vector explicitly depends on the measurements, which renders this processor nonlinear

$$w_{\mathrm{MVD}}(\mathbf{r},\mathbf{r}_i;\omega) = \frac{\Sigma_{j=1}^{m}[K_{ij}(\omega)]^{-1}w_B(\mathbf{r},\mathbf{r}_j;\omega)}{\Sigma_{i=1}^{m}\Sigma_{j=1}^{m}w_B{*}(\mathbf{r},\mathbf{r}_i;\omega)[K_{ij}(\omega)]^{-1}w_B(\mathbf{r},\mathbf{r}_j;\omega)}. \tag{A4}$$

Use of (A4) for the microphone weights in (A1) produces the MVD MFP output: $B_{\mathrm{MVD}}(\mathbf{r};\omega)$. In (A4), the $[\ ]^{-1}$ implies a matrix inverse which becomes problematic when $K_{ij}(\omega)$ does not have full rank. This problem occurs when $m>1$ but only one set of measurements $M_i(\omega)$ is used to construct $K_{ij}(\omega)$. Here, this matrix inversion problem was overcome by diagonal loading of $K_{ij}(\omega)$ (Baggeroer *et al.*, 1993) with 0.01 times the identity matrix δ_{ij}. The reasoning behind this choice is provided in Yönak and Dowling (1999).

For both processors, the extension from single to multiple frequencies is accomplished by incoherently averaging the narrow-band ambiguity functions

$$B(\mathbf{r}) = \frac{1}{N}\sum_{n=1}^{N} B(\mathbf{r};\omega_n), \tag{A5}$$

where the number of signal frequencies is N.

APPENDIX B: REPLICA FIELD MODELING

Matched-field processing requires calculated acoustic fields (replica fields) for source localization in known environments. Replica field models should be the Green's function solutions, $G(\mathbf{r},\mathbf{r}_i;\omega)$, to the Helmholtz equation at the frequencies of interest in the environment of interest. These Green's functions are folded into MFP through Eq. (A3). Here, three field models are defined for three acoustic environments: free space, a rectangular corner, and an infinite cylinder protruding from a infinite flat surface.

Using the coordinate system shown in Fig. 1, the free-space Green's function, G_o, between a hypothetical test source at $\mathbf{r}=(x,y,z)$ and a microphone location $\mathbf{r}_i=(x_i,y_i,z_i)$ will be

$$G_o(\mathbf{r}_i,\mathbf{r};\omega) = \frac{\exp\{ik\sqrt{(x_i-x)^2+(y_i-y)^2+(z_i-z)^2}\}}{4\pi\sqrt{(x_i-x)^2+(y_i-y)^2+(z_i-z)^2}}, \tag{B1}$$

where $k=\omega/c$ and c is the speed of sound.

In a rectangular corner, the method of images can be used to generate the replica field for a point source. Again, using the coordinate system of Fig. 1, consider the quarter-space acoustic domain defined by $x>0$ and $z>0$ bounded by flat, rigid reflecting surfaces at $y=0$, and $z=0$. The Green's function for this ideal rectangular corner, G_{rc}, is a sum of four free-space terms

$$\begin{aligned} G_{rc}(\mathbf{r}_i,\mathbf{r};\omega) = {} & G_o(\mathbf{r}_i,\mathbf{r},\omega) + G_o(\mathbf{r}_i,\mathbf{r}-2y\hat{y};\omega) \\ & + G_o(\mathbf{r}_i,\mathbf{r}-2z\hat{z};\omega) \\ & + G_o(\mathbf{r}_i,\mathbf{r}-2y\hat{y}-2z\hat{z};\omega), \end{aligned} \tag{B2}$$

where $\hat{x}$, $\hat{y}$, and $\hat{z}$ are unit vectors along the coordinate directions.

In the third geometry, a vertical cylinder of diameter $2a$ replaces the vertical reflecting surface. Although there is an exact Green's function for this situation (see Hanish, 1989), an approximate Green's function based on geometrical ray acoustics was constructed to reduce the computational expense of the MFP calculations. The ray theory provides a correction to the amplitude of the sound that reflects from the cylindrical surface (see Pierce, 1989). Thus, the following approximate form for the Green's function is obtained

$$\begin{aligned} G_{rc}(\mathbf{r}_i,\mathbf{r};\omega) = {} & G_o(\mathbf{r}_i,\mathbf{r};\omega) + G_o(\mathbf{r}_i,\mathbf{r}-2y\hat{y};\omega) \\ & + S_a\frac{\exp\{ik(R_a+l_a)\}}{4\pi R_a} \\ & + S_b\frac{\exp\{ik(R_b+l_b)\}}{4\pi R_b}. \end{aligned} \tag{B3}$$

Here, the first two terms are unaltered from (B2) because they represent the direct and the table-bounce paths that do not interact with the cylinder. The amplitude weightings (S_a,S_b) and the distances (R_a,R_b) in the third and fourth terms in (B3) are defined by

$$S_{a,b} = \left[\frac{R_{a,b}^2}{(R_{a,b}+l_{a,b})[R_{a,b}+l_{a,b}+2l_{a,b}(R_{a,b}/a)(\sin^2\phi_{a,b}\sec\vartheta+\cos^2\phi_{a,b}\cos\vartheta)]}\right]^{1/2}, \tag{B4}$$

$$R_{a,b} = \sqrt{(x-a\sin\theta_c)^2+(y-Y_{a,b})^2+(z-a\cos\theta_c)^2}, \tag{B5}$$

$$l_{a,b} = \sqrt{(x_i-a\sin\theta_c)^2+(y_i-y_{a,b})^2+(z_i-a\cos\theta_c)^2}, \tag{B6}$$

$$\phi_a=\frac{\pi}{2}-\tan^{-1}\left\{\frac{y_i-y}{\sqrt{(x_i-x)^2+(z_i-z)^2}}\right\}, \tag{B7}$$

$$\phi_b=\frac{\pi}{2}-\tan^{-1}\left\{\frac{y_i+y}{\sqrt{(x_i-x)^2+(z_i-z)^2}}\right\}, \tag{B8}$$

$$Y_a=y+\frac{(y_i-y)\sqrt{(x-a\sin\theta_c)^2+(z-a\cos\theta_c)^2}}{\sqrt{(x-a\sin\theta_c)^2+(z-a\cos\theta_c)^2}+\sqrt{(x_i-a\sin\theta_c)^2+(z_i-a\cos\theta_c)^2}}, \tag{B9}$$

$$Y_b=-y+\frac{(y_i+y)\sqrt{(x-a\sin\theta_c)^2+(z-a\sin\theta_c)^2}}{\sqrt{(x-a\sin\theta_c)^2+(z-a\cos\theta_c)^2}+\sqrt{(x_i-a\sin\theta_c)^2+(z_i-a\cos\theta_c)^2}}, \tag{B10}$$

$$\vartheta=\cos^{-1}\left\{\frac{\sin\theta_c(x-a\sin\theta_c)+\cos\theta_c(z-a\cos\theta_c)}{\sqrt{(x-a\sin\theta_c)^2+(z-a\cos\theta_c)^2}}\right\}, \tag{B11}$$

$$\theta_c=\tan^{-1}\left\{\frac{x+x_i}{z+z_i}\right\}. \tag{B12}$$

In spite of its trigonometric and algebraic complexity, the strength of this field model is its computational simplicity. Its weakness is its inability to handle shadowing and diffraction.

Baggeroer, A. B., Kuperman, W. A., and Schmidt, H. (**1988**). ''Matched Field Processing: Source Localization in Correlated Noise as an Optimum Parameter Estimation Problem,'' J. Acoust. Soc. Am. **83**, 571–587.

Baggeroer, A. B., Kuperman, W. A., and Mikhalevsky, P. N. (**1993**). ''An Overview of Matched Field Methods in Ocean Acoustics,'' IEEE J. Ocean. Eng. **18**, 401–424.

Bell, A. G. (**1880**). ''On the Production and Reproduction of Sound by Light,'' Am. J. Sci. **20**, 305–324.

Bell, A. G. (**1881**). ''Upon the Production of Sound by Radiant Energy,'' Philos. Mag. **11**, 510–528.

Brassington, D. J. (**1982**). ''Photo-acoustic Detection and Ranging—A New Technique for the Remote Detection of Gases,'' J. Phys. D **15**, 219–228.

Bucker, H. P. (**1976**). ''Use of Calculated Sound Fields and Matched-Field Detection to Locate Sound Sources in Shallow Water,'' J. Acoust. Soc. Am. **59**, 368–373.

Claspy, P.C. (**1977**). ''Infrared Optoacoustic Spectroscopy and Detection,'' in *Optoacoustic Spectroscopy and Detection*, edited by Yoh-Han Pao (Academic, New York), pp. 133–166.

Dewey, Jr., C. F. (**1974**). ''Opto-Acoustic Spectroscopy,'' Opt. Eng. **13**, 483–488.

Dewey, Jr., C. F., Kamm, R. D., and Hackett, C. E. (**1973**). ''Acoustic Amplifier for Detection of Atmospheric Pollutants,'' Appl. Phys. Lett. **23**, 633–635.

Goldan, P. D., and Goto, K. (**1974**). ''An Acoustically Resonant System for Detection of Low-Level Infrared Absorption in Atmospheric Pollutants,'' J. Appl. Phys. **45**, 4350–4355.

Hanish, S. (**1989**). *A Treatise on Acoustic Radiation* (Naval Research Laboratory, Washington, D.C.), Chap. 3.

Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H. (**1994**). *Computational Ocean Acoustics* (American Institute of Physics, New York), Chap. 10.

Kreuzer, L. B. (**1977**). ''The Physics of Signal Generation and Detection,'' in *Optoacoustic Spectroscopy and Detection*, edited by Yoh-Han Pao (Academic, New York), pp. 1–25.

Kreuzer, L. B., and Patel, C. K. N. (**1971**). ''Nitric Oxide Air Pollution: Detection by Optoacoustic Spectroscopy,'' Science **173**, 45–47.

Lymann, J. L., Quigley, G. P., and Judd, O. P. (**1986**). ''Single-infrared-frequency studies of multiple photon excitation and dissociation of polyatomic molecules,'' in *Multiple-Photon Excitation and Dissociation of Polyatomic Molecules*, edited by C. D. Cantrell (Springer, Berlin), pp. 9–94.

McRae, T. G. (**1994**). ''Photo Acoustic Leak Location and Alarm on the Assembly Line,'' Mater. Eval. **52**, 1186–1190.

McRae, T. G., and Dewey, A. H. (**1992**). ''Photo-acoustic leak detection system and method,'' US Patent No. 5,161,408.

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (Princeton, NJ), Chap. 12.

Pierce, A. (**1989**). *Acoustics* (American Institute of Physics, New York), p. 417.

Rosencwaig, A. (**1980**). *Photoacoustics and Photoacoustic Spectroscopy* (Wiley, New York).

Tolstoy, A. (**1993**). *Matched Field Processing For Underwater Acoustics* (World Scientific, River Edge, NJ).

Westwood, E. K. (**1992**). ''Broadband Matched Field Source Localization,'' J. Acoust. Soc. Am. **91**, 2777–2789.

Yönak, S. H. (**2000**). ''Characterization and Matched-Field Processing Localization of Photoacoustic Signals,'' Ph.D. thesis, University of Michigan, Ann Arbor, MI.

Yönak, S. H., and Dowling, D. R. (**1999**). ''Photoacoustic Leak Detection and Localization,'' J. Acoust. Soc. Am. **105**, 2685–2694.

Yönak, S. H., and Dowling, D. R. (**2001**). ''Multiple Microphone Photoacoustic Leak Detection and Localization System and Method,'' US Patent No. 6,227,036.

Ziomek, L. J. (**1995**). *Fundamentals of Acoustic Field Theory and Space-Time Signal Processing* (CRC Press, Boca Raton, FL).

Pyrotechnic shock response predictions combining statistical energy analysis and local random phase reconstruction[a)]

E. Bodin[b)] and B. Brévart
Alcatel Space Industries, 26, avenue J.-F. Champollion, B.P. 1187, 31037 Toulouse, Cédex 1, France

P. Wagstaff
"Génie des systèmes mécaniques" department, Université de Technologie de Compiègne, 60206 Compiègne Cédex, France

G. Borello
Interac, 10, impasse Borde Basse, 31240 L'Union, France

(Received 22 January 2001; accepted for publication 31 October 2001)

Numerous pyrotechnic devices are used on satellites to separate structural subsystems, deploy appendages, and activate on-board operating subsystems. The firing of these pyrotechnic mechanisms leads to severe impulsive loads which could sometimes lead to failures in electronic systems. The objective of the present investigation is to assess the relevance of a method combining deterministic calculations and statistical energy analysis to predict the time overall shock environment of electronic equipment components. The methods are applied to the low- and high-frequency ranges, respectively, which may be defined using a modal parameter based on the effective transmissibility. Initially, in order to address the problem of the low-frequency content of the mechanical shock pulse, the linear dynamic response of the equipment was calculated using direct time integration of a finite element model of the structure. The inputs in the form of the accelerations measured in all three directions at each of the four bolted interfaces were injected into the model. The high-frequency content of the shock response is taken into account by considering the intrinsic dynamic filtering of the equipment. This frequency filter magnitude is extrapolated from the transfer function given by statistical energy analysis between the different imposed accelerations and the response accelerations. Their associated phases are synthesized by considering pseudo-modal phase variations around the group velocity of the structural flexural waves. Combining the effects of the high-frequency filter outputs and the low-frequency finite element calculations yields good predictions of the equipment shock time response over the whole frequency range of interest. [DOI: 10.1121/1.1446050]

PACS numbers: 43.40.Jc [PJR]

I. INTRODUCTION

Past experience has shown that the pyrotechnic shocks generated in the environment of satellites have been the origin of numerous failures,[1] and even today the effects of these shocks on the electronic equipment are difficult to master. Their severity is increasing and this has led to a significant raise in the shock specifications of all major constructors. It is therefore becoming necessary to characterize the nature and the level of the equipment shock response. Because some components such as relays are sensitive to the maximum peak acceleration, it is of interest to predict the shock time history, using the accelerations measured at the equipment–satellite interface.

The potential of standard finite element modeling to predict the low-pass-filtered transient behavior of the electronic equipment has been demonstrated.[2] When the acceleration is known exactly in all three directions at each of the equipment interfaces, direct time integration yields good results up to the mid-frequency range. At higher frequencies, the inaccuracies of the finite element model (FEM), with regard to components, connectors, bolts, and other details, result in larger discrepancies. The objective of the present investigation is to assess the relevance of a method combining low-frequency deterministic calculations with statistical energy analysis at high frequencies to predict the overall shock environment of the electronic equipment in the time domain.

At the end of the 1980s, it was attempted[3] to apply a nonstationary SEA concept to transient problems, but the method revealed limitations since energetic balance is not verified for each frequency band. The originality of the present work consists in implementing a high-frequency time signal synthesis, based on stationary SEA and a pseudo-random phase calculation. The approach is similar to the one followed by Dalton,[4,5] but it is based on a distinct phase synthesis. His method uses a virtual mode synthesis. It supposes that modes are distributed over frequency according to the modal density estimate and that these modes collectively produce the frequency response envelope at each resonant frequency. The virtual mode coefficients are obtained by comparing the FRF magnitude of the virtual system with that obtained by SEA. The present approach is simpler and more

[a)]Part of this work was presented in "Pyrotechnic shock response predictions combining direct time FEM and SEA based frequency filters," Proceedings of Novem2000, Lyon, France, August 2000.
[b)]Electronic mail: emmanuel.bodin@space.alcatel.fr

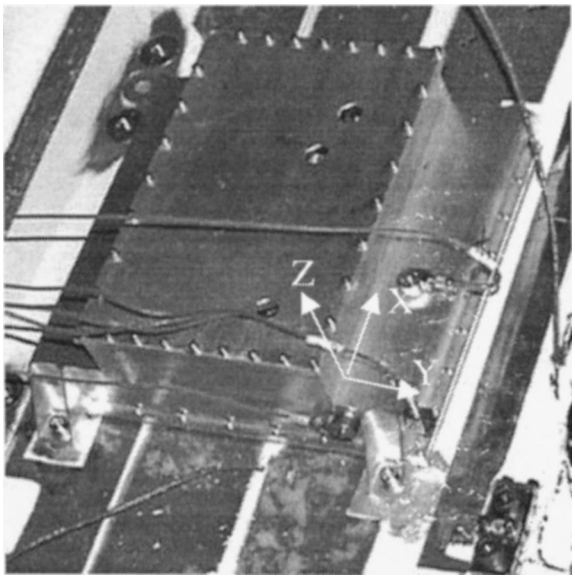


FIG. 1. Test equipment.

realistic. It uses a random phase concept justified by the high recovering of the modes at high frequency.

Initially the limits of the low- and high-frequency domains have to be defined. In order to address the low-frequency content of the mechanical shock pulse, the linear dynamic response of the equipment was calculated using direct time integration of a finite element model of the system. On the other hand, the high-frequency content of the shock response was taken into account by considering the intrinsic dynamic filtering of the equipment. The amplitude of the transmissibilities is obtained by SEA and the phase is obtained by the novel method of synthesizing. Combining the effects of the high-frequency filter outputs and the low-frequency finite element calculations yields good predictions of the equipment shock time response over the whole frequency range of interest.

II. TEST DESCRIPTION AND RESULTS

The prediction method presented in this study is illustrated through the simulation of the transient response of an equipment printed circuit board (PCB) to a mechanical shock. The equipment (Fig. 1) consists of a rectangular aluminum box containing three circuit boards which are attached to the box with screws around their periphery. The box is bolted at its four corners to the test block. Impacts of the type metal on metal delivered by the test facility shown in Fig. 2 are used. They are assumed to be representative of the effects of the pyrotechnic shock excitation suffered by the equipment during its mission.[2] The shock response spectra (SRS) of the input and PCB accelerations are represented in Fig. 3. The SRS function represents the maximum response acceleration to the shock of a one degree of freedom (DOF) system at each represented frequency (Fig. 4). The signal at equipment interface is broadband and reaches 7000 g, as shown by its SRS at very high frequencies, while the SRS on the PCB reveals that its time acceleration maximum is only 2000 g. Two distinct frequency ranges may be distinguished. Below 6 kHz, the excitation is amplified by the modal response characteristics of the equipment and, above this frequency, the input is highly filtered. The time frequency analysis (Fig. 5) of the PCB response illustrates this behavior. The spectrogram has been calculated by applying Hanning windows on 2048 point blocks of the signal sampled at 50 kHz. The low-frequency components LF, il-

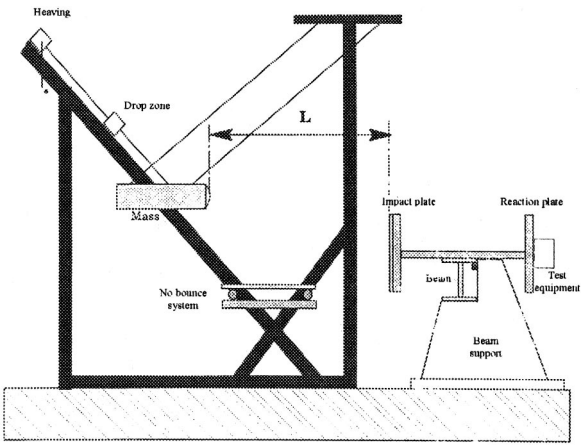


FIG. 2. Metal on metal shocks machine.

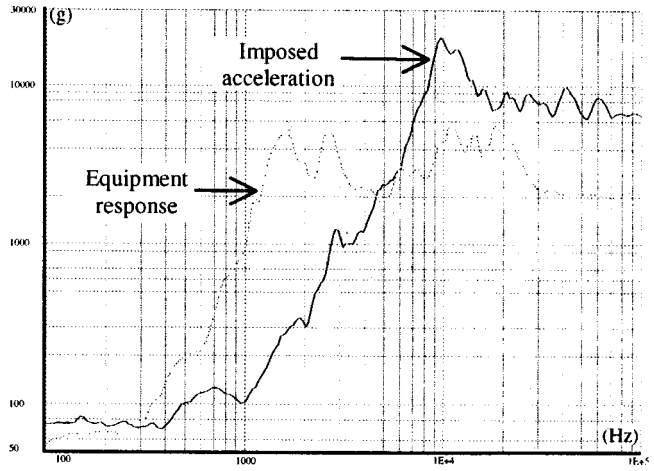


FIG. 3. Typical shock on electronic equipment.

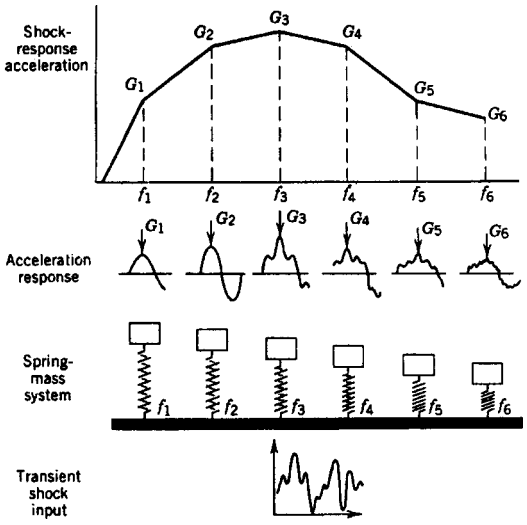


FIG. 4. Definition of SRS.

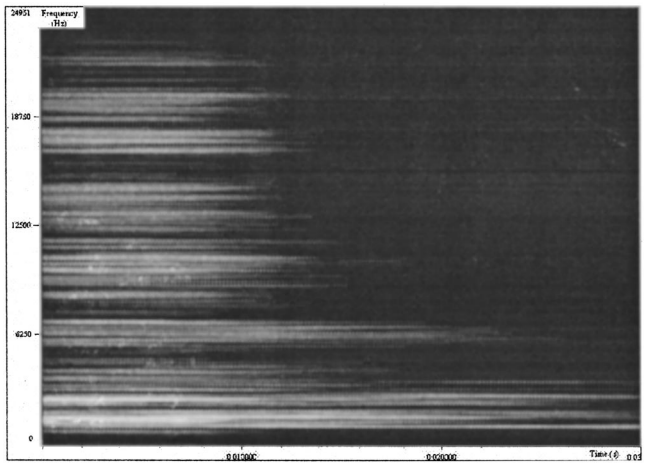


FIG. 5. Equipment acceleration spectrogram.

lustrated by the persistent horizontal lines from approximately 1 to 6.5 kHz, are the result of the equipment modal response, generated by the initial wave system and its successive reflections. The HF components are illustrated by short horizontal lines which represent the result of the dynamic filtering of the imposed accelerations.

III. TRANSITION BETWEEN THE FREQUENCY DOMAINS

The transition frequency between low- and high-frequency ranges may be predicted using the effective transmissibility parameter.[6] This modal parameter indicates the level of transmitted acceleration due to global modal behavior, from a junction to any equipment point (on the PCB, for instance). It may therefore be used to define the maximum frequency of the LF domain and the minimum frequency of the HF domain.

To extract this parameter, a modal analysis of the equipment finite element model (FEM) was performed. Clamped boundary conditions were applied at the interface with the fixing points. Thus the effects of phase differences in the injected signals on the system modal behavior are neglected. The system could then be reduced to a first substructure consisting of the interface degrees of freedom DOF and a second represented by a set of DOFs on the equipment model. The equation of motion leads to Eq. (1) if damping is neglected:

$$\begin{bmatrix} M_{jj} & M_{ji} \\ M_{ij} & M_{ii} \end{bmatrix} \begin{bmatrix} \ddot{u}_j \\ \ddot{u}_i \end{bmatrix} + \begin{bmatrix} K_{jj} & K_{ji} \\ K_{ij} & K_{ii} \end{bmatrix} \begin{bmatrix} u_j \\ u_i \end{bmatrix} = \begin{bmatrix} R_j \\ 0 \end{bmatrix}. \tag{1}$$

$[M]$ and $[K]$ are mass and stiffness matrices, respectively. u_j represents the displacements imposed at the junction and u_i represents the internal displacements of the forced harmonic response. R_j represents the force reactions at the junctions. The eigenmodes of both substructures are introduced and transformations of Eq. (1) lead to the dynamic transmissibility matrix which links u_j and u_i:

$$\{u_j\} = [\tilde{T}_{ij}(\omega)]\{u_j\}, \tag{2}$$

$$[\tilde{T}_{ij}(\omega)] = \sum_{k=1}^{m} T_k(\omega) \cdot \tilde{T}_{ij,k}, \tag{3}$$

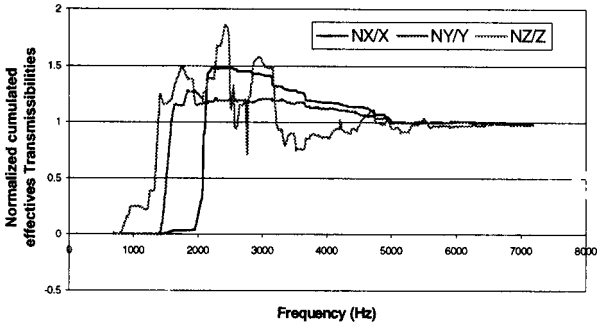


FIG. 6. Cumulated effective transmissibilities, between the interface and the PBC.

where $T_k(\omega)$ represents the dynamic transmissibility factor of the mode k. The effective transmissibility matrix $\tilde{T}_{ij,k}$ of the mode k is given by

$$\tilde{T}_{ij,k} = \frac{[\Phi_{ik}].[L_{kj}]}{m_k}, \tag{4}$$

where $[\Phi_{ik}]$ symbolizes the matrix of truncated eigenmodes for clamped interface boundary conditions and m_k represents the generalized mass of the mode k. Equation (5) defines the participation factor L_{kj} of the mode k where $[S_{ij}] = -[K_{ii}] \times [K_{ij}]^{-1}$ is the junction static mode matrix:

$$L_{kj} = [\Phi_{ik}]([M_{ii}][S_{ij}] + [M_{ij}]). \tag{5}$$

The expression of the dynamic transmissibility matrix reveals the contribution of each mode to the equipment response. This contribution is made up of the dynamic transmissibility factor and a term which characterizes the mode. Moreover, if an infinite modal base is considered, the sum of all of the modal effective transmissibility matrices is exactly equal to the junction static mode matrix:

$$\sum_{k=1}^{+\infty} [\tilde{T}_{ij,k}] = \sum_{k=1}^{+\infty} \frac{[\Phi_{ik}][L_{kj}]}{m_k} = [S_{ij}]. \tag{6}$$

A relative transmissibility matrix can be defined for a truncated base of n modes as follows:

$$\lim_{n \to +\infty} \left(\sum_{k=1}^{n} [\tilde{T}_{ij,k}] \right) = [S_{ij}]. \tag{7}$$

Since the interface is isostatic, the $[S_{ij}]$ matrix includes the rigid modes of the equipment model. This occurs when the rows, corresponding to an internal DOF, and the columns, corresponding to a junction DOF, simultaneously correspond to the same direction with the other elements at zero. The summation of the effective transmissibilities between the interface DOF (X, Y, Z translations) and a PCB DOF (X, Y, Z translations, respectively) was calculated and is presented in Fig. 6. It converges to around 97% of the static mode transmissibility around 6 kHz. Thus, above this frequency, no more acceleration is transmitted to the observed point of the PCB due to global modal behavior. It should be noted that the junction static mode matrix has no information concerning coupling relations between the axes because of the decoupling introduced by the modal decomposition.

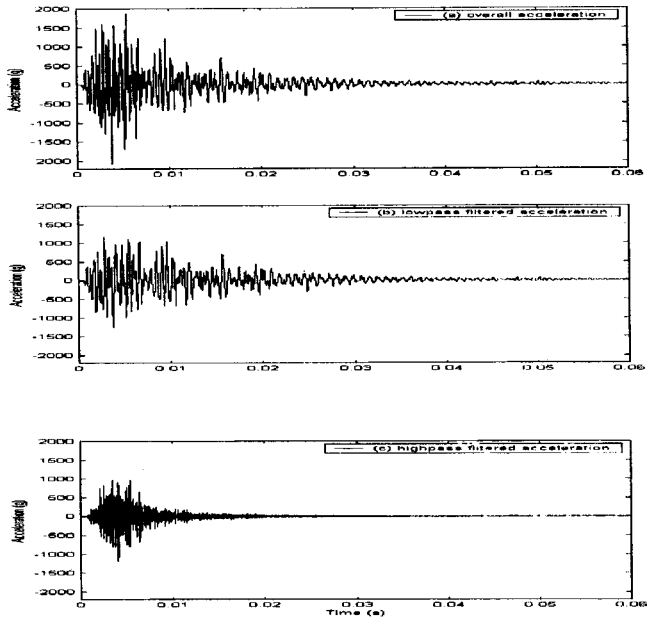


FIG. 7. (a) Equipment response, (b) low pass filtered, (c) high pass filtered.

This cutoff frequency is used to filter the time recording of the PCB response. Figure 7(b) shows that the time history of its LF components has a maximum of around 1240 g, approximately half as much as the unfiltered response [Fig. 7(a)]. Above 6.5 kHz significant spectral components result in peaks up to 1180 g [Fig. 7(c)] in the unfiltered time domain signal [see Fig. 7(a)]. This example shows that a prediction based uniquely on the modal response will result in errors of around 50% in the maximum of the time domain response. Inversely an approach based uniquely on the HF response would minimize the strong modal response of the equipment. Only a time domain approach combining the transient low-frequency response of the equipment with the HF components of the filtered input acceleration can reproduce the entire severity of the shock applied to an electronic component.

IV. CALCULATION OF THE LF PART OF THE RESPONSE

The method used to predict the LF part of the response of the electronic equipment used a direct time response resolution. It permits multiple tri-axial inputs to be applied and the simulation of a damped input wave propagation. The finite element model of the equipment was adapted to take into account the totality of the vibration waves that could propagate in the 0–10-kHz band which is more than sufficient. An implicit form was adopted to solve the system's time response because it is adapted to fast calculations of the LF response to small deformations. The shock is modeled by a distribution of forced accelerations, using Lagrangian multipliers, applied at the nodes corresponding to the lower plane of the four feet of the equipment casing. At each of the feet the accelerations are considered to be perfectly homogeneous assuming that the surface of contact interfacing with the structure is infinitely rigid. The results are calculated using NASTRAN.

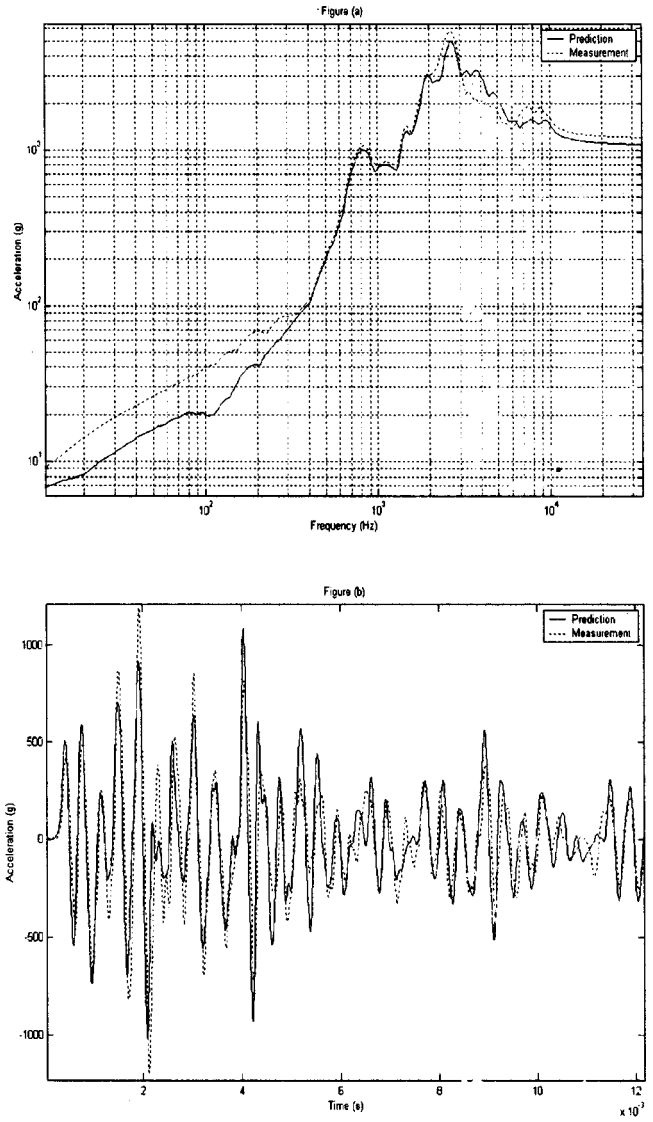


FIG. 8. Equipment response calculation (Pamshock), (a) SRS and (b) time record, in the *z* direction.

For numerous shocks, all of around 2200 g, the quasi-totality of the response of the equipment is contained in the 0- to 10-kHz range. In these cases 98% of the maximum of the time response is reconstituted. Figure 7 compares the time domain response and the predicted values [Fig. 8(b)] and the measured and predicted shock response spectra (SRS) [Fig. 8(a)]. The transient response calculated for the equipment casing corresponds very closely to the measurement, leading to a good correlation with all the derived functions such as the SRS. The maximum in the time domain is reconstituted at 90% of its true value and the results obtained for the PCB are also satisfactory. In this case where the majority of the acceleration inputs has been filtered by the dynamic response of the system, the deterministic approach gives good predictions.

The results confirm the pertinence of the time domain approach to predict the response of the electronic equipment from acceleration inputs if the latter are available. In cases where the excitation produces high-frequency components in the response it is necessary to determine the contribution of

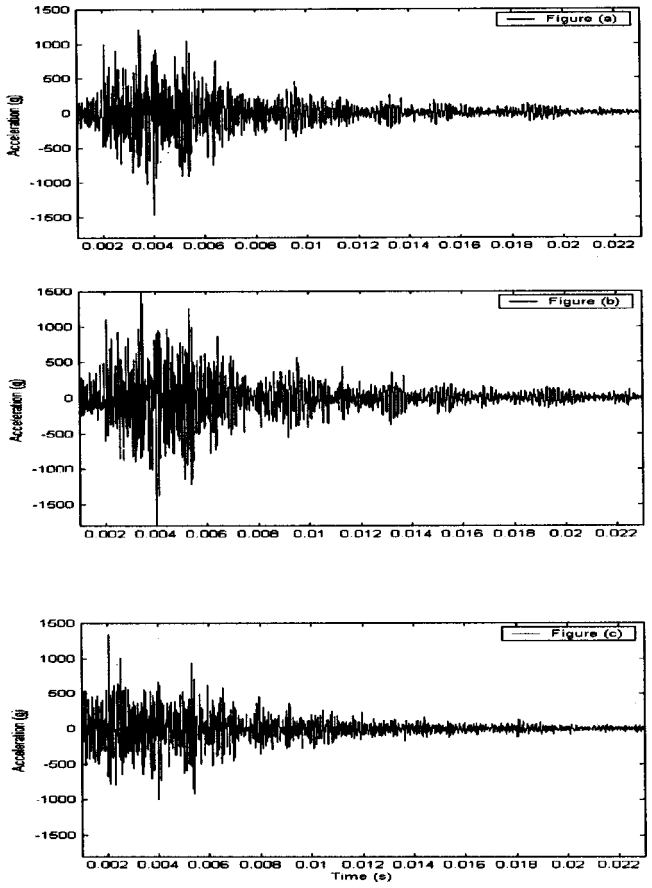


FIG. 9. Filtered equipment acceleration, (a) measured, (b) reconstructed using modified magnitude and exact phase, (c) reconstructed using exact magnitude and modified phase.

these HF components to the response of the equipment in the time domain.

V. SYNTHESIS OF THE HF COMPONENTS OF THE RESPONSE

R. H. Lyon[7] has suggested that the phase of the frequency response function (FRF) between two points A and B of a structure is more important than the modulus in determining the response at B due to a transient excitation at A. This is verified in our application for the FRF between an imposed acceleration of the equipment and the acceleration response, particularly at high frequencies. Figure 9(a) shows the signal measured on the equipment and filtered at 6.5 kHz. In Figs. 9(b) and (c) the synthesized results for the response to the same signal are presented using a modified transmissibility, first with degraded values of the modulus and then with degraded values of the phase. The degradation is simulated by using one point of the FRF instead of 150 and by applying linear interpolations between the retained samples. The errors introduced in the modulus have little influence on the response, suggesting that SEA predictions of the average modulus of transmissibility for each frequency band can lead to a valid prediction of the HF response of the equipment. However, larger discrepancies are introduced through the phase modifications, implying that a precise method to predict the phase is required.

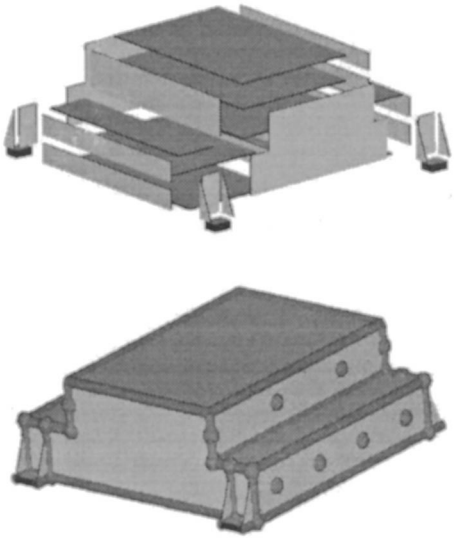


FIG. 10. SEA model.

A. Calculation of the transmissibility magnitude

To calculate the transmissibilities between the forced and response accelerations of the equipment, a SEA model was formulated using AutoSEA 2.[8] The model presented in Fig. 10 was divided in 15 subsystems and is valid for frequencies above 3 kHz. The traditional experimental analysis using a shock hammer is limited in its frequency range (0–9 kHz) due to the difficulty of obtaining sufficient force input energy at the higher frequencies. Nevertheless, coupling loss factors between internal elements have been identified experimentally from 0 up to 9 kHz and the representations of the analytical connections have been adjusted from the extrapolation of these estimations at higher frequencies.

In the case where the four tri-axial inputs applied to the model are known, it would be possible to use 12 transmissibility functions between the desired response point on the PCB normal to its surface and each of the excitations. Nevertheless, the shock has a predominant response in the *z* direction, permitting the study to be limited to this one direction. Additionally, the symmetry of the equipment permits a

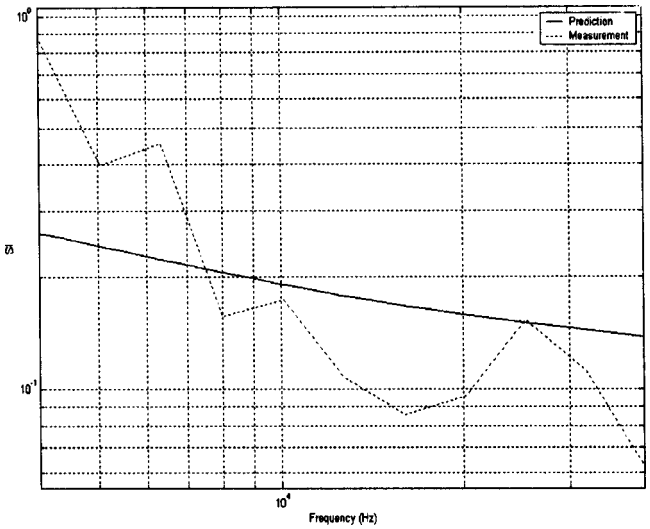


FIG. 11. Measured and predicted (SEA) transmissibility magnitude.

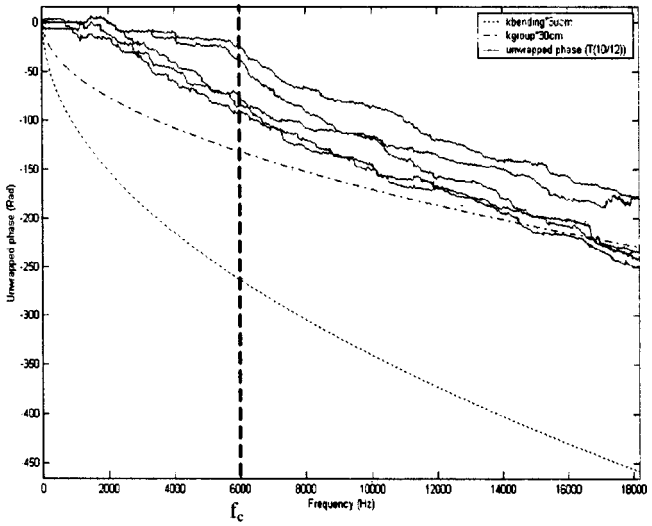


FIG. 12. Transmissibility unwrapped phase (one interface/box).

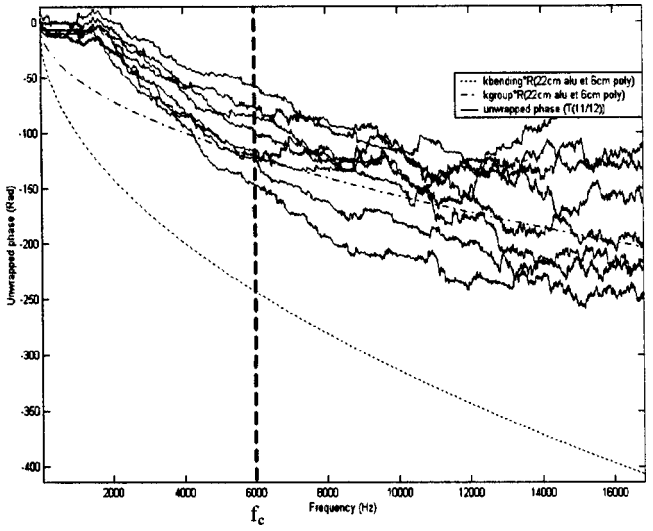


FIG. 14. Transmissibility unwrapped phase (one interface/PCB).

single transmissibility function modulus (Fig. 11) to be used between each of the four feet and the PCB. Using loss factors of 10% for the aluminum components of the casing and 20% for the PCB in polyimide, the differences between the predicted and measured values are inferior to 6 dB between 6.5 and 50 kHz. The model can therefore be considered satisfactory. Initially these loss factors appear to be excessively high, but it should be noted that they are derived from very high amplitude shock measurements. For high levels of vibrations, in transient or in random excitation, the higher the level of excitation the higher the loss factors. Additionally, connectors, screws, and bolts which participate to a high degree in the increased loss factors are not included in the model. Nevertheless, despite the fact that these factors seem to be reasonable for certain subsystems such as the PCBs, they are more difficult to justify for the main structure of the aluminum casing of the equipment. This indicates that certain of the coupling coefficients are poorly estimated. This is certainly true at the level of the feet because of the small dimensions of these elements.

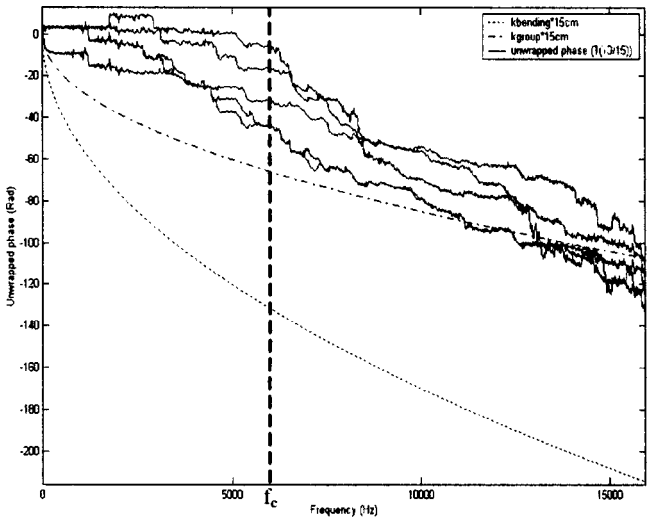


FIG. 13. Transmissibility unwrapped phase (one interface/box).

B. Local random phase reconstruction

The significance of the phase of the transmissibility between two points has already been demonstrated. This leads to the conclusion that each of the transmissibility function magnitudes determined by SEA between the feet and the PCB in the *z* direction should be accompanied by an accurate synthesized phase relationship. Numerous shock hammer tests on the equipment interfaces show that the response function at B due to a forced acceleration at a point A seems to obey certain physical rules. Figure 12 shows the unwrapped phases of the transmisibilities between an acceleration imposed at one of the four feet and the acceleration at a point on the casing for five averaged shock inputs. Figure 13 shows the same type of result for a source input closer to the response point and Fig. 14 shows the result for a response on the PCB. The progressive unwrapped phases for these different measurements do not correspond to the phase lag which one would associate with the slowest structural wave, i.e., the bending wave propagating in a semi infinite system (1-mm-thick aluminum sheet) over a distance *r* separating the points A and B.

In fact, the measured unwrapped phases seem to correspond to the lag associated with the group speed or the propagation of a packet of waves between two points. Here the product $(k_{\text{groupe}}.r)$ is defined as the group delay of the bending waves associated with the different materials traversed and the average distances of propagation. The similitude is valid for frequencies above the transition frequency *fc*, where the structural modes are no longer predominant in the response. The energy appears to propagate between two points of observation at the group velocity of the flexural wave.

Below this frequency, the clearly defined changes in phase associated with the normal modes can be observed varying around $\pm\pi$, following the constant values. The duration of these constant phase responses differs for the transmissibility between the different types of pairs of points as can be seen in Fig. 14 between a foot and a point on the PCB or between the same foot and a point on the casing as shown

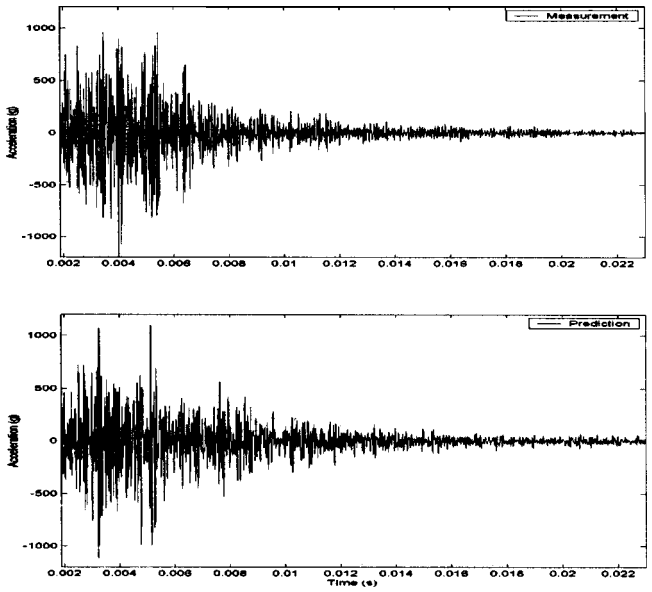


FIG. 15. HF response of the PCB.

in Figs. 12 and 13 (for 0–4 kHz). Effectively the modes of the PCB card, which is more flexible than the casing, participate more in the response amplitude at lower frequencies than the casing. Much of the energy is localized in the PCB.

Additionally the different unwrapped phase responses which, at high frequencies have characteristics close to $(k_{\text{groupe.}}r)$, are not exact functions of $\sqrt{\omega}$ as the group delay should be. There are numerous oscillations that can be related to the modal density. The elevated number of modes per frequency band implies a large number of phase changes in the transmissibility and, at high frequencies, the overlapping modes produce arbitrary oscillations in the response which are random in nature. Thus in the HF band the phase oscillates continuously between $[(k_{\text{groupe.}}r)-\pi]$ and $[(k_{\text{groupe.}}r)+\pi]$.

To synthesize a realistic representation of the phase measured between the excitation of a foot A and its response at a point B it is interesting to adopt a pseudo-random phase

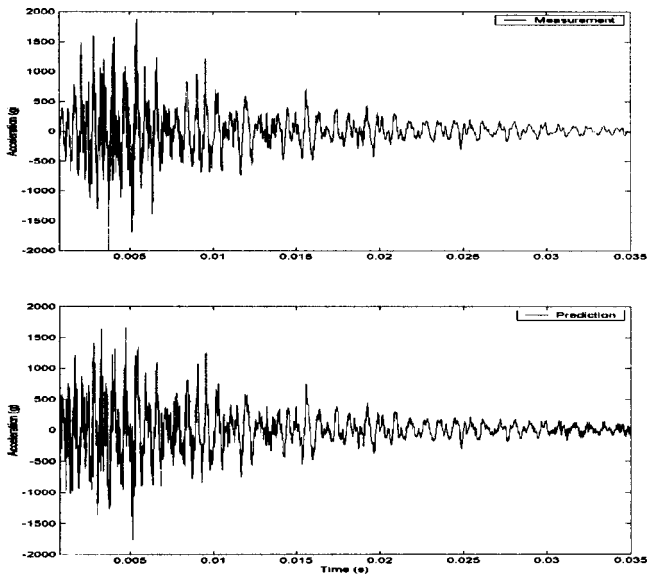


FIG. 16. Overall time response of the PCB.

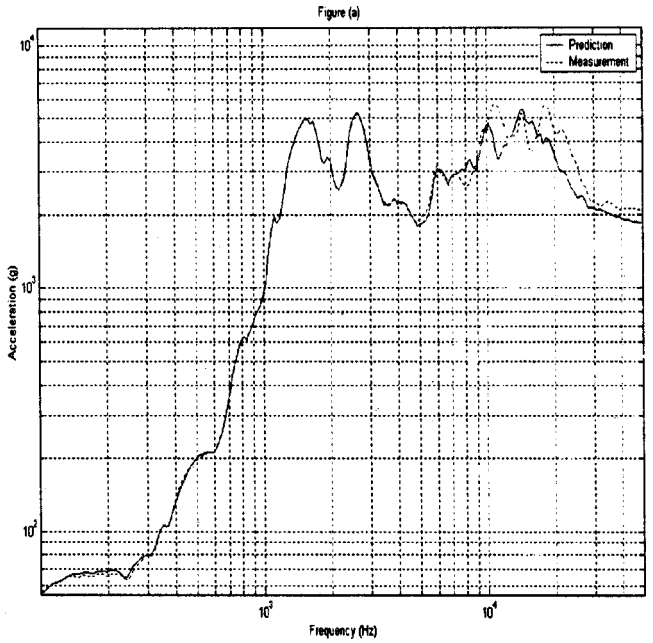


FIG. 17. Overall response (SRS) of the PCB.

variation reflecting the presence of numerous local modes of the assembly of plates constituting the casing at the previously determined group speed. The phase is synthesized by using a linear distribution of the number of modes in each of the frequency bands and associating a random variation in the interval $[-\pi, \pi]$ following a uniform distribution.

VI. SYNTHESIS OF THE TIME OVERALL SHOCK RESPONSE

At this stage, the HF part of the equipment response can be predicted. The moduli of the HF transmissibilities in third octave bands predicted by SEA and the associated wrapped phases have been defined from f_0 (the center frequency of the lowest frequency band integrating the transition frequency) up to the Nyquist frequency of the accelerations measured during the shock experiments. They are resampled at the sampling frequency of the accelerations and then recombined. The transmissibilities constructed between Ai (Ai is the excitation vector at the foot 1, 2, 3, or 4, in the z direction) and Bz are then multiplied by the accelerations measured at Ai, at the base of the system during the shock. Zero padding is applied between 0 and f_0 and the signal is antisymmetrized to obtain a real function. The resulting acceleration is passed through a high-pass FIR filter, whose cutoff frequency is the transition frequency (6.5 kHz) of the LF and HF domains. The aim is to keep the right frequency components and to solve the f_0 discontinuity problem. The obtained HF content of the response is presented in Fig. 15. The maximum acceleration is correctly reproduced, in a proportion of 87%.

The overall response of the PCB is given by adding up the LF and HF accelerations. Figure 16 presents the response in the time domain, and Fig. 17 presents the corresponding SRS. In order to emphasize the significance of the HF prediction, the LF part emerges from the LF filtering of the measurement and the HF one is synthesized according to the

methodology presented in Sec. V. An excellent correlation between measured and predicted time signals is obtained, as the recovered maximum reaches 90% of its true value. A good consistency of the SRS with the measurement follows, with a difference less than 6 dB.

VII. CONCLUSIONS

This study presents a method of predicting the response of an electronic equipment assembly to high-frequency shocks. This method combines the use of deterministic calculations and statistical energy analysis. The FEM provides the LF content of the acceleration and SEA, coupled to a local random phase reconstruction concept, provides the HF content. The application of such an approach yields promising results and illustrates its capacity for modeling the effects of pyrotechnic shocks.

[1] C. J. Moening, ''View of the world of pyrotechnic shock,'' in *The Shock and Vibration Bulletin*, no. 56, Part 3, August 1986, p. 3/28.

[2] E. Bodin, B. Brévart, and T. Youssefi, ''Predictions of an electronic equipment response to a pyrotechnic shock excitation,'' in *Launch Vehicle Vibrations Proceedings*, Toulouse, 14–16 December 1999, pp. 397–404.

[3] D. Lednik and R. J. Pinnigton, ''Prediction of high frequency vibrations induced by pyrotechnic shock sources of separation,'' ISVR Contract Report no. 90/11 (1989).

[4] E. C. Dalton, ''Ballistic shock response analysis by an extension of SEA,'' Cumming Research Park, Huntsville, AL (1990).

[5] E. C. Dalton, ''High frequency shock prediction short course at Military Technology, Inc.,'' Huntsville, 22–24 June 1999.

[6] A. Girard, ''Réponse des structures à un environnment basse-fréquence,'' Note Technique no. 116, CNES (1985).

[7] R. H. Lyon, *Machinery Noise and Diagnostics* (Butterworths, Massachusetts Institute of Technology, 1987).

[8] AutoSEA 2 User Guide, Vibro-acoustic Sciences Inc., 1999, www.vasci-.com, AutoSEA-X 2 User Guide, InterAC Sarl, 1999, www.interac.fr.

Evaluation of a noise reduction system for the assessment of click-evoked otoacoustic emissions[a)]

Pascal Müller and Martin Kompis[b)]
Department of ENT, Head, Neck and Cranio-Maxillo-Facial Surgery, Inselspital, University of Berne, 3010 Bern, Switzerland

(Received 17 January 2002; revised 1 May 2002; accepted 1 May 2002)

A recently proposed noise reduction system intended to facilitate the assessment of click-evoked otoacoustic emission (CEOAE) in noisy environments [Comput. Biol. Med. **30**, 341 (2000)] is evaluated using 13 normally hearing ears and 9 ears with a sensorineural hearing loss. The noise reduction system is based on an adaptive noise canceller design using an additional noise-only reference microphone and intended to reduce externally generated noise. The system is tested in quiet and at different levels of white noise. The three main design parameters of the noise reduction system (adaptation time constant, length of the adaptive filter, and position of the noise reference microphone) are varied systematically in different experiments. With the noise reduction system active, CEOAE can be assessed correctly at noise levels which are 5 to 9 dB higher than without the noise reduction system. For the range of adaptation time constants considered (65.6 to 656 ms), no statistically significant effect on the amount of noise reduction is observed. Noise reduction is highest when the reference microphone is positioned close to the ear probe. Using this reference microphone position and adaptive filters of 6.56 ms in length, average noise reductions of 7.17 to 8.50 dB are achieved. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1488138]

PACS numbers: 43.50.Hg, 43.60.Lq, 43.64.Jb, 43.64.Yp [MRS]

I. INTRODUCTION

Otoacoustic emission (OAE) are low-level acoustic signals emitted by the normally hearing ear (Kemp, 1978; Wilson, 1980). OAE are routinely assessed in clinical practice to differentiate between ears with no or a relatively mild hearing loss, and ears with a more pronounced hearing loss and consequently no measurable OAE. The attractiveness of OAE measurements for clinical purposes stems from its noninvasiveness, the short measuring time, and the objective nature of its results.

Several methods to assess OAE have been proposed (Probst *et al.*, 1991; Maat *et al.*, 2000). One method which is widely used in clinical practice and the method considered in this investigation is the assessment of click-evoked otoacoustic emission (CEOAE). To measure CEOAE, an ear probe containing a sensitive microphone and a miniature loudspeaker is inserted into the outer ear canal. Acoustic stimuli (pulse trains, i.e., series of ''clicks'') are emitted through the loudspeaker of the ear probe and the resulting otoacoustic emissions originating from the inner ear are measured by the probe microphone (Kemp, 1978; Probst *et al.*, 1991).

The sound level of CEOAE is low, typically around 20 dB sound-pressure level (SPL) (Jannsen, 2001). As a consequence, unrelated external noise can interfere considerably with the assessment of CEOAE (Rhoades *et al.*, 1998). This is particularly restricting, as CEOAE are used in neonatal screening programs and are often measured in less than optimal acoustic conditions. For neonatal units, which are typical sites for CEOAE screening, ambient noise levels of 50 to 75 dB(A) have been reported (Jacobson and Jacobson, 1994; Kemp and Ryan, 1991). In contrast, noise levels between 30 and 55 dB(A) have been recommended for OAE assessment (Jacobson and Jacobson, 1994; Rhoades *et al.*, 1998).

Several noise sources that influence the measurement of OAE have been identified (Popelka *et al.*, 1998). Noise sources can be subdivided into two groups: external noises (such as noise from unrelated technical equipment or sounds from other persons in the same room) and noises generated by the subject, such as heavy breathing or coughing. Such internal noises tend to cause excessively high noise levels in the ear canal, which usually result in considerable amounts of nonlinearity already at the level of the probe microphone. Reduction of such noise from internal sources is very difficult (Kompis *et al.*, 2000) and beyond the scope of this investigation. However, based on the data provided by Jacobson and Jacobson (1994), it can be expected that suppression of external noises alone by several dB should be useful and clinically relevant.

To minimize the detrimental effect of external noise on CEOAE measurements, several measures have been identified and/or proposed. Proper sealing of the ear probe in the ear canal has been found to be essential (Kemp and Ryan, 1991). To improve the signal-to-noise ratio (SNR), typically, several hundred individual responses to acoustic stimuli are averaged. To further optimize SNR in the final data set, individual responses or time segments with low SNR (detected by their temporarily increased total signal variance) can be rejected automatically by the OAE system. Such an algorithm is implemented, e.g., in the ILO88 (Otodynamics Ltd., Hatfield, UK) OAE system (Kemp *et al.*, 1990). However, as

[a)]Portions of this work were presented in ''Ein neues Verfahren zur Unterdrückung von Störgeräuschen bei der Messung otoakustischer Emissionen,'' Proceedings of the *Deutsche Gesellschaft für Audiologie, Dritte Jahrestagung,* Hannover, Germany, March 2000.

[b)]Electronic mail: martin.kompis@insel.ch

 0001-4966/2002/112(1)/164/8/$19.00

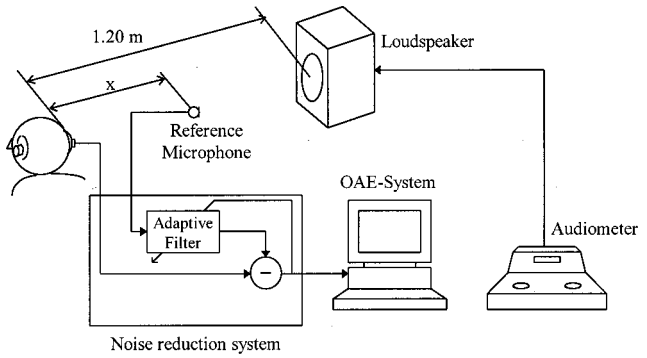


FIG. 1. Schematic drawing of the experimental setting, showing a volunteer with the ear probe in his right ear and an earplug in his contralateral ear. A schematic drawing of the investigated noise reduction system is shown in the lower left corner.

an increasing fraction of the individual time segments is excluded at higher noise levels, CEOAE measurements can take an unacceptably long time or can even be blocked completely under noisy conditions.

For these reasons, there seems to be a need for additional methods to reduce external noise in CEOAE measurements. Recently, a promising noise reduction system based on the principle of the adaptive noise canceller (Widrow *et al.*, 1975) has been proposed and implemented in real time (Kompis *et al.*, 2000). Similar adaptive noise reduction methods have already been applied successfully in other areas in acoustics, e.g., to auscultation in noise (Patel *et al.*, 1998) or to noise reduction for hearing aids (Vanden Berghe and Wouters, 1998). To demonstrate the potential of this concept for the assessment of CEOAE, results from measurements from a single subject and using a single set of design parameters have been reported earlier (Kompis *et al.*, 2000). It was shown that the proposed system is indeed able to reduce external noise by 7–8 dB and that this noise reduction occurs predominantly in the clinically important frequency band between 700 and 3000 Hz. However, these preliminary results do not show to what extent the proposed method could be useful for patients other than the one subject considered and how the amount of noise reduction could be optimized by a proper choice of design parameters of the noise reduction system. The present study aims to close these gaps.

II. MATERIALS AND METHODS

A. Description of the noise reduction system

A block diagram of the noise reduction system under evaluation is shown in the lower left part of Fig. 1. The system is based on an adaptive noise canceller, as proposed by Widrow *et al.* (1975). An external reference microphone is mounted in the vicinity of the ear probe but outside of the ear canal in order to record external noise, but not the otoacoustic emissions. External noise which enters the ear canal and appears as part of the ear probe microphone signal usually differs in phase and in amplitude from the noise recorded by the reference microphone. An adaptive filter accounts for these differences in such a way that the filtered reference microphone signal can be subtracted from the noisy probe microphone signal to increase its SNR. The noise reduction system has been implemented in real time in a stand-alone unit built around an ADSP-2181 digital signal processor. The realization of the system used in this study is described in detail elsewhere (Kompis *et al.*, 2000) and only the data relevant to the current investigation are repeated here.

The length of the adaptive filter can be chosen between $N=1$ and 180 coefficients in steps of 1, corresponding to filter lengths of up to $F=6.56$ ms at the sampling rate of the system of $f_{\text{sample}}=27.42$ kHz. Filter coefficients are updated continuously using a normalized least-mean-squares algorithm. The dimensionless adaptation constant α can be varied by the user in the range of 0.00 to 0.05 in steps of 0.005. The adaptation time constant T_{adapt} of the system can be calculated as

$$T_{\text{adapt}}=\frac{N}{2\alpha\cdot f_{\text{sample}}}. \tag{1}$$

Longer adaptation time constants cause less accurate tracking of rapid changes in the acoustic environment, but yield smaller residual errors of the adaptive filter in slowly varying settings (Widrow *et al.*, 1975). The signal delay introduced by the noise reduction system is 1.2 ms. The unit interfaces seamlessly with the ILO92 OAE system (Otodynamics, Hatfield, UK) used in this study. A rotary switch enables the user to by-pass the noise reduction system without the need to change any connections between measurements.

B. Experimental setting

Numerous methods of measuring the attenuation of hearing protection devices, which reduce noise in the ear canal, have been proposed (Berger, 1986). However, as the investigated system reduces noise in the probe microphone signal rather than in the ear canal, the experimental setting shown in Fig. 1 was used. Subjects were seated in a medium-sized (49 m^3), moderately reverberant room (average reverberation time 0.34 s, 250–8000 Hz) with an ear probe from an ILO92 CEOAE system in the test ear. The test ear was oriented toward a loudspeaker (Phonak LAB 50, Phonak, Staefa, Switzerland) positioned at the height of the head of the subject at a distance of 1.2 m. This distance was chosen to yield an estimated direct-to-reverberant ratio of 1. The direct-to-reverberant ratio is the ratio between the variance of the signal due to the directly incident portion of the acoustic signal and reverberation. The omnidirectional reference microphone (ECM-T140, Sony, Japan) was held in position by a microphone stand at different distances x from the test ear. The reference microphone was positioned along an imaginary line through the test ear and the center of the loudspeaker. $x=0$ was approximated by positioning the reference microphone less than 1 cm away from the OAE ear probe. For $x<0$, the reference microphone was placed at the side of the head contralateral to the test ear.

Noise was generated by a GSI61 clinical audiometer (Grason-Stadler, Milford, NH) operating in the white-noise setting. A typical spectrum, as measured in the ear canal by

the ear probe microphone, is shown in Fig. 3. Noise levels were measured with a handheld level meter (type 116, Norsonic AS, Tranby, Norway; A-weighted level, time constant ''slow''). The level of the background noise at the position of the test ear, almost entirely generated by the cooling fan of the computer of the OAE-system, was 36 dB (A).

OAE measurements were performed using the ILO92 OAE hardware and the ILO88 CEOAE software (both from Otodynamics, Hatfield, UK). For each single-CEOAE measurement, responses to 260 stimuli were averaged into two buffers, A and B. The time window for the analysis was 2.5–19.3 ms if the noise reduction system was by-passed, and 3.7–20.5 ms if the noise reduction was active. In this way, the length of the time window remains 16.8 ms, and the shift of the time window by 1.2 ms compensates for the delay introduced by the noise reduction system. The cutoff value for noise, above which individual measurements are rejected by the ILO88 software as being too noisy for analysis, was set to its maximal level of 54.9 dB. Stimulus level was between 75 and 80 dB for all measurement. Acoustic input to the contralateral ear was reduced by earplugs (E.A.R. classic, Aearo Ltd., Stockport, UK) in order to minimize any effects on CEOAEs in the test ear (Mott *et al.*, 1989; Collet *et al.*, 1990). The specified average attenuation of these earplugs is 24.6 to 41.6 dB in the range of 250 to 4000 Hz.

C. Experimental protocol

Experiments were performed with normally hearing subjects and subjects with a substantial, purely sensorineural hearing loss. After giving informed consent, each volunteer underwent an examination including otoscopy, pure-tone audiometry (air and bone conduction thresholds) tympanometry, and the measurement of CEOAE in quiet on both ears. On the basis of this examination, it was determined for each ear separately whether (a) it qualified for inclusion in the study as a normal hearing ear; (b) it qualified for inclusion in the study as a hearing impaired ear; or (c) it did not qualify for the inclusion in the study.

To qualify as normally hearing, there had to be (i) no history of ear diseases and (ii) no evidence of any ear disease during the otologic examination; (iii) a pure-tone hearing thresholds of 20 dB hearing level (HL) or better between 250 and 8000 Hz with air–bone gaps of 10 dB or less; (iv) normal tympanometry; and (v) normal CEOAE in quiet, i.e., with a response level of at least 5 dB SPL and a wave reproducibility above 50% (Schmutziger *et al.*, 1998).

To qualify for the inclusion as hearing impaired, requirements were (i) a pure sensorineural hearing loss with an air–bone gap of 10 dB or less; (ii) pure-tone thresholds of 30 dB HL or more in the entire frequency range 2000 to 8000 Hz; (iii) normal tympanograms; and (iv) no measurable CEOAE in quiet (i.e., wave reproducibility 30% or less).

Three different experiments, numbered I, II, and III, each addressing different aspects of the validation of the noise reduction system, were performed.

The objective of experiment I was to reproduce the results of the above-mentioned earlier report using a single volunteer (Kompis *et al.*, 2000) and to ascertain whether the proposed noise reduction system performs satisfactorily for *all* tested ears from a larger group of volunteers, both with normal hearing and sensorineural hearing losses according to the above criteria. The reference microphone was placed in the middle between test ear and loudspeaker ($x=0.60$ m), corresponding to the setting used by Kompis *et al.* (2000). The adaptation constant of the noise reduction system was set to 0.005 and the filter length F was set to its maximal value of 6.56 ms ($N=180$ filter coefficients). CEOAE were measured both in quiet and at noise levels between 47 and 69 dB in steps of 1 dB. For each noise level, one measurement with the noise reduction system by-passed and one measurement with the noise reduction system active was performed. If, in a given test condition, test time exceeded 120 s due to numerous rejected responses at high noise levels, the measurement was aborted. For each completed measurement, response level, wave reproducibility, noise level (as assessed by the OAE system in the input signal), and test time were noted and used as raw data to estimate the amount of noise reduction. Figure 2 shows examples of two complete data sets, one for a normally hearing ear and one for a hearing-impaired ear.

The objective of experiment II was to assess the influence of the adaptation constant α. CEOAE were measured with the noise reduction by-passed and with the noise reduction active using three different values for α (0.005, 0.02, and 0.05), corresponding to adaptation time constants T_{adapt} of 656, 164, and 65.5 ms, respectively, at the maximal filter length of $F=6.56$ ms used. Data for $\alpha=0.005$ were actually already collected in experiment I, so these measurements were not repeated. The position of the reference microphone was the same as in experiment I.

The objective of experiment III was to investigate the influence of the length of the adaptive filter and the microphone position. Eight different positions of the reference microphone were tested ($x=-0.40$, -0.20, 0.00, 0.20, 0.40, 0.60, 0.80, and 1.00 m). For each microphone position, CEOAE were measured with the noise reduction system active and three different lengths of the adaptive filter ($N=45$, 90, and 180 coefficients), corresponding to filter lengths of 1.64, 3.28, and 6.56 ms, respectively. The adaptation constant was set to $\alpha=0.01$ in experiment III.

Several methods to estimate the amount of noise reduction from the raw data were considered. Because of the intended clinical application of the evaluated system, only measures which are based on the availability of useful clinical data at higher noise levels are used. For experiments with normally hearing ears, two different outcome measures, Rnh and Tnh, are used. Figure 2 illustrates how these measures are computed. The difference between the highest noise levels in dB, at which CEOAE can be assessed within 120 s with the noise reduction active and the noise reduction by-passed, respectively, is labeled Tnh (''time, normal hearing''). The difference between the highest noise levels (in dB), at which wave reproducibility is 50% or higher with the noise reduction by-passed and the noise reduction system active, respectively, is labeled Rnh (''reproducibility, normal hearing''). For hearing-impaired ears, wave reproducibility never reached 50%. Therefore, for this group, only the dif-

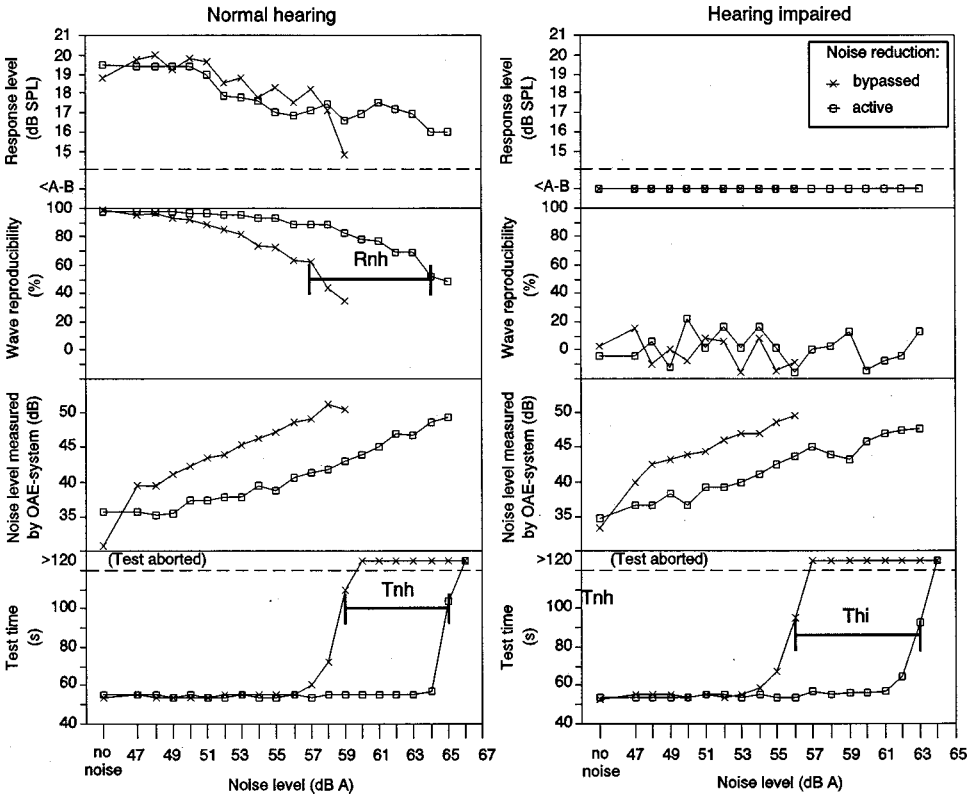


FIG. 2. Examples of two sets of raw data from experiment I. Left: normally hearing ear, right: hearing-impaired ear. The solid horizontal lines labeled Rnh, Tnh, and Thi illustrate how the three outcome measures Rnh, Tnh, and Thi are computed.

ference between the highest noise levels at which test time for CEOAE does not exceed 120 s with the noise reduction by-passed and the noise reduction active, respectively, is used and labeled Thi (''time, hearing impaired'').

The experimental protocol was approved by the Ethical Committee of the Canton Berne.

D. Subjects

Eight adult volunteers (5 women, 3 men, age 22 to 37, mean 25) with subjectively normal hearing and 6 adult volunteers with a substantial bilateral hearing loss (3 women, 3 men, age 26 to 74, mean 60) participated in this study.

From the 16 ears tested in the normal hearing group, 3 ears in 3 different subjects (2 left, 1 right) did not meet the criteria listed in Sec. II C, leaving 13 normally hearing ears for this investigation. From the 12 ears tested in the hearing-impaired group, 3 ears in 3 different subjects (2 left, 1 right) did not meet all of the corresponding criteria, leaving 9 hearing-impaired ears.

All 13 normal-hearing ears and all 9 hearing-impaired ears were tested in experiment I. Participation in experiments II and III required the OAE probe to be fitted tightly in the ear canal for several hours. In clinical practice, OAE measurements usually take only a few minutes. The design of the ear probe cannot be expected to be optimized for the extended time periods required in experiments II and III, and varying amounts of discomfort as well as progressively poor mechanical fit was observed in several ears. As a consequence, experiments II and III were completed only on a subset of 6 normally hearing ears (4 right, 2 left) in 5 different subjects and 6 hearing-impaired ears (2 right, 4 left) in 6 different subjects in which the above problems either did not occur or were minimal. Ears that were included in experiments II and III did not, as a group, differ from the other ears in terms of the noise reduction measured in experiment I (Mann–Whitney U-test).

III. RESULTS

A. Experiment I

In experiment I, measurements were performed with all 13 normally hearing ears and 9 hearing-impaired ears. Figure 2 shows examples of two sets of results from experiment I, one for a normally hearing ear and one for a hearing-impaired ear. In the two cases shown, it was not possible to measure CEOAE at any noise levels above 66 or 64 dB (A), respectively, even with the noise reduction system active, and no data points are shown above these values.

For the normally hearing ear, both response level and wave reproducibility decrease with increasing levels of external noise. However, at high noise levels the values remain higher with the noise reduction system active. The highest noise level at which wave reproducibility is above 50% is 57 dB (A) without noise reduction and 64 dB (A) with noise reduction, corresponding to an Rnh of 7 dB. The noise level, as measured by the OAE system as its input, decreases by approximately 7 dB with the noise reduction system active. The highest noise level at which a CEOAE measurement can be completed in less than 120 s is 59 dB (A) without the noise reduction system and 65 dB (A) with the noise reduction system, corresponding to a Tnh of 6 dB.

For the hearing-impaired ear, no response level can be estimated by the OAE system (resulting in a ''<A–B'' message by the OAE system, where A and B are the two averaging buffers) and wave reproducibility clearly remains be-

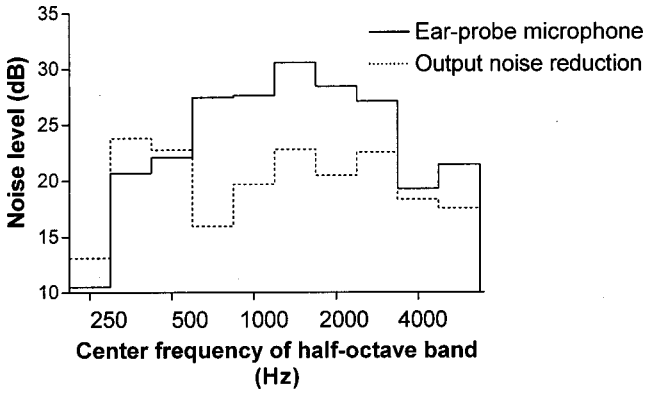


FIG. 3. Typical spectra of the noise signal as recorded by the ear-probe microphone (solid line) and at the output of the noise reduction system, averaged over half-octave bands. Noise reduction is substantial mainly between 700 and 3000 Hz. (Noise levels in dB *re*: arbitrary units).

low 30% for all conditions considered. The highest noise level at which CEOAE measurements can be completed in less than 120 s is 56 dB (A) without the noise reduction system and 63 dB (A) with the noise reduction system, corresponding to a Thi of 7 dB.

These two examples are typical in a number of aspects in all ears tested. For every ear in the study, the noise reduction system allowed a complete CEOAE assessment within 120 s at varying, but considerably higher noise levels than with the system by-passed. Wave reproducibility remained below 30% for every hearing-impaired ear in every test condition. For every normally hearing ear, wave reproducibility decreased with increasing noise levels, but remained above 50% at considerably higher noise levels when the noise reduction system was active. Presumably because of system noise of the noise reduction system, in quiet, noise levels (as assessed by the OAE system) were generally slightly higher with the noise reduction system active. However, the absolute level of this additional noise is too low to interfere with CEOAE measurements or with the interpretation of their results. Occasionally, especially after reinsertion of the ear probe after a break in a long series of measurements, small (i.e., usually less than 1 dB) changes in response levels were observed. They probably reflect small changes in the position of the earplug. However, over all measurements, no systematic trend towards either higher or lower response levels with the noise reduction system active was observed.

Figure 3 shows typical spectra of the noise signal as measured in the ear canal and at the output of the noise reduction system, using half-octave bands. A substantial reduction of the noise can be observed mainly between 700 and 3000 Hz. Above 3000 Hz, the noise level already decreases in the ear canal, presumably due to the damping of the tissue and the ear probe.

The leftmost panel in Fig. 4 (labeled $\alpha=0.005$) shows a synopsis of the amounts of noise reduction in the 13 normally hearing ears and the 9 hearing-impaired ears. Noise reductions between 5 and 9 dB were obtained in each case. For each of the three measures (Rnh, Tnh, Thi), noise reduction is statistically highly significantly different from 0 ($p<0.002$, nonparametric sign test). Average values and standard deviations for Rnh, Tnh, and Thi are listed in Table I. The average noise reduction assessed by the outcome measure Rnh is higher than Tnh by 1.31 dB. This difference is statistically significant ($p<0.05$, nonparametric sign test). There is no statistically significant different between normally hearing ears and hearing-impaired ears, as assessed by Tnh and Thi.

B. Experiment II

In experiment II, the influence of the adaptation time constant α was investigated. Figure 4 shows the results of this experiment, along with the results from experiment I in the leftmost panel ($\alpha=0.005$) for comparison. For the range of values considered for the adaptation constant α between 0.005 and 0.05, there is no statistically significant difference in the amount of noise reduction seen for Rnh, Tnh, or Thi (Wilcoxon signed rank test, $p>0.05$). However, the tendency for the two measures considered for normally hearing ears, Rnh and Tnh, which was already observed in experiment I, is accentuated at higher values of α.

C. Experiment III

In experiment III, the influence of the length of the adaptive filter and the position of the reference microphone was investigated. Figure 5 shows the results of experiment III. At the two microphone positions with negative x values (reference microphone positioned at the contralateral side of the test ear), the average noise reduction is low, typically between 0 and 2 dB. For microphone positions between the test ear and the loudspeaker ($x=0$ to 1 m), noise reduction is considerably higher and both filter length F and position of the reference microphone x as well as interaction between those two parameters do have a statistically significant effect

TABLE I. Synopsis of average noise reduction (± standard deviation) for four different test conditions (x = distance between ear probe and reference microphone, F=length of adaptive filter).

Noise reduction measure	$x=0.60$ m (Experiment I)	$x=0.00$ m (Experiment III)		
	F=6.56 ms	F=1.64 ms	F=3.28 ms	F=6.56 ms
Rnh	7.69 (±1.18)dB	5.83 (±1.72)dB	7.50 (±1.38)dB	8.50 (±0.84)dB
Tnh	6.38 (±1.12)dB	7.00 (±1.90) db	7.50 (±1.52)dB	7.17 (±1.47)dB
Thi	7.00 (±1.22)dB	7.67 (±1.75)dB	8.67 (±2.16)dB	8.33 (±1.75)dB

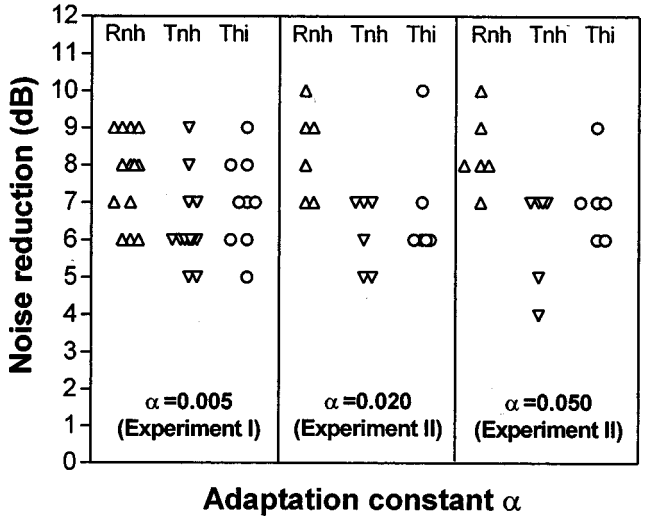


FIG. 4. Noise reduction as a function of the adaptation constant α. Left panel: results from experiment I. Middle and right panels: results from experiment II. Each symbol denotes a result from one set of measurements for one ear. See the text for explanation of measures Rnh, Tnh, and Thi; see Fig. 2 for a graphical representation.

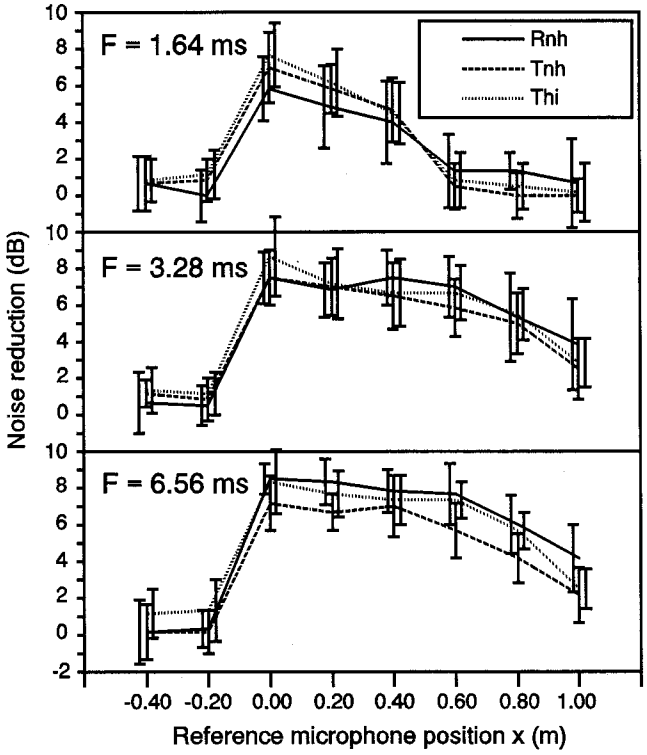


FIG. 5. Noise reduction as a function of microphone position x and the length of the adaptive filter F (experiment III). Error bars denote standard deviation, x denotes the distance between the ear probe and the reference microphone.

on the amount of noise reduction ($p<0.0001$, Wald Test, Brunner and Langer, 1999). The shortest filter length of F=1.64 ms, which corresponds to a distance of approximately 0.55 m for a speed of 340 m/s in air, yields only small noise reductions once x exceeds this distance. For microphone positions between $x=0$ and $x=0.40$ m, noise reduction increases slightly with filter length F, but the differences are relatively small even for an increase in filter length F by a factor of 4 (cf. Table I). Noise reduction tends to be highest for $x=0.00$. In Table I, average values and standard deviations of Rnh, Tnh, and Thi for this microphone position are listed and compared to the results from experiment I. Despite the tendency towards higher noise reductions for longer filter lengths, at $x=0$ only the difference in Rnh between the longest and the shortest filter length is statistically significant ($p<0.05$, two-sided Wilcoxon signed rank test, paired).

IV. DISCUSSION

The most important finding of this study is that the proposed noise reduction system is indeed able to provide considerable amounts of noise reduction for OAE measurements in all tested ears, regardless of the outcome measure (Rnh, Tnh, or Thi) employed or the hearing ability of the tested ears. In no case did the noise reduction system alter the clinically important parameter ''wave reproducibility'' in a way which could lead to a misclassification of a normally hearing ear as hearing impaired, or vice versa. The amount of noise reduction was in the range from 4 to 10 dB, depending on the chosen outcome measure and adaptation constant α (Fig. 4). This order of magnitude should be sufficient to make the noise reduction system clinically useful (Jacobson and Jacobson, 1994). The observed improvements are in the range which can be expected for this type of noise reduction systems. It confirms the preliminary findings reported earlier (Kompis *et al.*, 2000) and agrees with findings for similar systems for hearing aid applications (Kompis and Dillier, 2001b). For these systems, it has been shown that room reverberation is a major limiting factor in realistic acoustic environments (Kompis and Dillier, 2001b).

No detrimental effect of any practical importance of the noise reduction system on the CEOAE measurements or its results has been observed. This holds true for measurements in noise and in quiet. The noise reduction system did not require any attention nor any adjustments by the user other than the changes of the parameters necessitated by the experimental protocol. If the reference microphone were included in an existing part of the OAE system, e.g., the external side of the OAE ear probe, the entire noise reduction system could be implemented in a way that would be ''invisible'' to the user while still providing a substantial benefit in terms of noise reduction.

Our original idea was to mount the reference microphone permanently close to the user interface or the computer of the OAE system, normally at some distance from the test ear, where it could be well protected mechanically and would not add any size, weight, or complexity to the ear probe. For this reason, the reference microphone was positioned at $x=0.60$ m in the preliminary study and in experiments I and II. However, experiment III shows clear disadvantages of this approach. To be able to cancel the nonreverberant part of an external noise, the distance between the test ear and the reference microphone must not exceed the filter length F multiplied by the speed of sound. An additional problem arises if the reference microphone happens to be farther away from the external noise source than the test ear (cf. Fig. 4, $x=-0.20$ and -0.40 m). In this

case, noise reduction is minimal for two reasons. The directly incident (nonreverberant) portion of the noise signal cannot be reduced properly by the adaptive filter, as the corresponding reference signal arrives too late with respect to the noise in the probe microphone signal. In addition, if the head of the subject is placed between the ear probe and the reference microphone in a reverberant room, the correlation between the two signals becomes small above approximately 1500 Hz (Lindevald and Benade, 1986). A part of this problem can be solved by delaying the probe microphone signal by F/2 before the output of the adaptive filter is subtracted and by simultaneously increasing filter length. However, signal delay and system complexity are increased.

Results of experiment III suggest that it may be more favorable to mount the reference microphone at the external side of the ear probe. This corresponds to $x=0.00$ m in experiment III. In this position, the proper choice of the filter length F becomes less critical, and no delays need to be introduced to account for different positions of external noise sources. In addition, presumably because the reverberant parts of the external noise signal are more similar if the probe microphone and the reference microphone are close to each other, noise reduction is increased by 0.81 dB (Rnh), 0.79 dB (Tnh), and 1.33 dB (Thi), respectively, when compared to the $x=0.60$ m reference microphone position (cf. Table I). Using these optimized design parameters ($x=0.00$ m, F=6.56 ms), average noise reductions were 7.17 dB (Tnh) to 8.50 dB (Rnh) (cf. Table I).

The specific amount of noise reduction obtained in this study may have been influenced by the specific positioning of the noise source and the resulting direct-to-reverberant ratio. As the correlation of the reverberant portions of the noise signal increases with decreasing distances between the probe microphone and the reference microphone, the influence is expected to be smallest for the position $x=0$ of the reference microphone. Even in a diffuse noise field, where no substantial amounts of noise cancellation can be expected if the two microphones are remote to each other, at $x=0$ the signals can still be expected to be correlated and thus the investigated system can be expected to provide some noise reduction.

The impact of the size of the adaptation constant α is small for values between 0.005 and 0.05 considered in experiment II. The increasing differences in noise reduction between the measures Tnh and Rnh, however, may be an indication of artifacts due to the adaptive filter. Such artifacts may indeed become larger when filter coefficients fluctuate more in the adapted state of the filter for higher values of α. This would also explain the relatively small but statistically significant difference between noise reduction in terms of Tnh and Rnh, already found in experiment I. Notable misadjustment of the filter (and, as a consequence, artifacts) usually appears immediately after signal segments with high amplitudes in the noisy input signal, which in this application are most notably the clicks emitted by the OAE system. Such artifacts can potentially increase overall noise level (as detected by the OAE system and used to reject measurements as too noisy) and/or alter the reproducibility of the time segments following each click. From measurements in quiet and in noise, we know that the contribution to the overall noise level detected by the OAE system is low when compared to the external noise at noise levels over 47 dB(A) (Fig. 2). However, values for Rnh and Tnh tend to drift apart for higher values of α, i.e., when more artifacts due to filter misadjustments can be expected, suggesting that reproducibility may indeed be affected at higher values of α. Consequently, we feel that Tnh is a more appropriate measure to define noise reduction than Rnh. To reduce the probability that such artifacts influence the results of CEOAE measurements, it seems reasonable to keep adaptations constants small, i.e., on the order of magnitude $\alpha \approx 0.005$ for the given filter length.

The influence of acoustic parameters of the experimental setting on the amount of noise reduction which can be obtained by noise reductions systems using adaptive filters is known to be complex (Kompis and Dillier, 2001a, b) and not expected to be understood on the basis of the above experiments alone. Nevertheless, the results of experiment I indicate that the system is able to reduce noticeable amounts of both direct and reverberant parts of external noise. Results of experiment III suggest that it may be more important to keep the distance between the ear probe and the reference microphone short than to keep the direct-to-reverberant ratio at the reference microphone high, e.g., by placing it close to a known noise source such as a noisy air-conditioning system.

Numerous interesting aspects are beyond the scope of this investigation. Further experiments would be needed to assess the influence of acoustical parameters of the environment, such as noise spectrum, reverberation time of the room, or the number and position of the external noise sources on the noise reduction system. Other projects of substantial clinical interest could be to determine the amount of noise reduction which can be expected in CEOAE tests in neonates (as opposed to adults in the current study) and to develop and test similar noise reduction techniques for distortion product otoacoustic emissions (DPOAE).

V. SUMMARY

A real-time noise reduction system (Kompis *et al.*, 2000), based on the principle of an adaptive noise canceller (Widrow *et al.*, 1975) intended to reduce externally generated noise is shown to provide significant amounts of noise reduction on the order of magnitude of 5 to 9 dB when measuring CEOAE in noisy environments. Results of experiment I, in which 13 normally hearing and 9 hearing-impaired ears are studied, indicates that the noise reduction system does not alter the clinically important parameter wave reproducibility in a way that might cause normally hearing ears to be misclassified as hearing impaired, or vice versa. Results from experiment III suggest that the amount of noise reduction can be maximized by mounting the required reference microphones on the outside of the ear probe. As the noise reduction system does not require any attention or adjustments by the user, it might be realized as an ''invisible'' but clinically useful addition to commercial CEOAE systems.

ACKNOWLEDGMENTS

We would like to thank S. Müller from the Department of Mathematical Statistics from the University of Berne for the major part of the statistical analysis of our data, and Mrs. E. Clamann for her help in preparing this text.

Berger, E. H. (**1986**). ''Methods of measuring the attenuation of hearing protection devices,'' J. Acoust. Soc. Am. **79**, 1655–1687.

Brunner, E., and Langer, F. (**1999**). ''Nichtparametrische Analyse longitudinaler Daten [Non-parametric analysis of longitudinal data]. Oldenburg, Munich (Germany).

Collet, L., Kemp, D. T., Veuillet, E., Duclaux, R., Moulin, A., and Morgon, A. (**1990**). ''Effect of contralateral auditory stimuli on active cochlear micro-mechanical properties in human subjects,'' Hear. Res. **43**, 251–262.

Jacobson, J. T., and Johnson, C. A. (**1994**). ''The effects of noise in transient EOAE newborn hearing screening,'' Int. J. Pediatr. Otorhinolaryngol. **29**, 235–248.

Janssen, T. (**2001**). ''Otoakustische Emissionen (OAE) [Otoacoustic emissions (OAE)],'' in *Praxis der Audiometrie*, 8th ed., edited by E. Lehnhardt and R. Laszig (Thieme, Stuttgart), pp. 79–107.

Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T., and Ryan, S. (**1991**). ''Otoacoustic emissions tests in neonatal screening programs,'' Acta Oto-Laryngol., Suppl. **482**, 73–84.

Kemp, D. T., Ryan, S., and Bray, P. (**1990**). ''A guide to the effective use of otoacoustic emissions,'' Ear Hear. **11**, 93–105.

Kompis, M., and Dillier, N. (**2001a**). ''Performance of an adaptive beamforming noise reduction scheme for hearing aid applications. I. Prediction of the signal-to-noise-ratio improvement,'' J. Acoust. Soc. Am. **109**, 1123–1133.

Kompis, M., and Dillier, N. (**2001b**). ''Performance of an adaptive beamforming noise reduction scheme for hearing aid applications. II. Experimental verification of the predictions,'' J. Acoust. Soc. Am. **109**, 1134–1143.

Kompis, M., Oberli, M., and Brugger, U. (**2000**). ''A novel real-time noise reduction system for the assessment of evoked otoacoustic emissions,'' Comput. Biol. Med. **30**, 341–354.

Lindevald, I. M., and Benade, A. H. (**1986**). ''Two-ear correlation in the statistical sound field of rooms,'' J. Acoust. Soc. Am. **80**, 661–664.

Maat, B., Wit, H. P., and van Dijk, P. (**2000**). ''Noise-evoked otoacoustic emissions in humans,'' J. Acoust. Soc. Am. **108**, 2272–2280.

Mott, J. B., Norton, S. J., Neely, S. T., and Warr, W. B. (**1989**). ''Changes in spontaneous otoacoustic emissions produced by acoustic stimulation of the contralateral ear,'' Hear. Res. **38**, 229–242.

Patel, S. B., Callahan, T. F., Callahan, M. G., Jones, J. T., Graber, G. P., Foster, K. S., Glifort, K., and Wodicka, G. R. (**1998**). ''An adaptive noise reduction stethoscope for auscultation in high noise environments,'' J. Acoust. Soc. Am. **103**, 2483–2491.

Popelka, G. R., Karzon, R. K., and Clary, R. A. (**1998**). ''Identification of noise sources that influence distortion product otoacoustic emission measurements in human neonates,'' Ear Hear. **19**, 319–328.

Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). ''A review of otoacoustic emissions,'' J. Acoust. Soc. Am. **89**, 2027–2067.

Rhoades, K., McPherson, B., Smyth, V., Kei, J., and Baglioni, A. (**1998**). ''Effects of background noise on click-evoked otoacoustic emissions,'' Ear Hear. **19**, 450–462.

Schmutziger, N., Hauser, R., and Probst, R. (**1998**). ''Zur diagnostischen Wertigkeit otoakustischer Emissionen—Teil 2. Klinische Einsatzfähigkeit der OAE [Diagnostic value of otoacoustic emissions—Part 2. Clinical usefulness of OAE],'' HNO **46**, 823–837.

Vanden Berghe, J., and Wouters, J. (**1998**). ''An adaptive noise canceller for hearing aids using two nearby microphones,'' J. Acoust. Soc. Am. **103**, 3621–3626.

Widrow, B., Glover, J. R., McColl, J. M., Kaunitz, J., Williams, C. S., Hearn, R. H., Zeidler, J. R., Dong, J. R., and Goodlin, R. C. (**1975**). ''Adaptive noise canceling: Principles and applications,'' Proc. IEEE **63**, 1692–1716.

Wilson, J. P. (**1980**). ''Evidence for a cochlear origin for acoustic re-emissions, threshold fine-structure and tonal tinnitus,'' Hear. Res. **2**, 233–252.

Ship recognition via its radiated sound: The fractal based approaches

Su Yang[a)]
State Key Laboratory of Modern Acoustics & Institute of Acoustics, Nanjing University, Nanjing 210093, People's Republic of China

Zhishun Li[b)]
National Key Laboratory of Underwater Information Processing and Control & College of Marine Engineering, Northwestern Polytechnical University, Xi'an 710072, People's Republic of China

Xinlong Wang[c)]
State Key Laboratory of Modern Acoustics & Institute of Acoustics, Nanjing University, Nanjing 210093, People's Republic of China

(Received 7 July 2000; revised 18 April 2002; accepted 18 April 2002)

Due to the complexity of its radiated sound, ship recognition is difficult. Fractal approaches are proposed in this study, including fractal Brownian motion based analysis, fractal dimension analysis, and wavelet analysis, to augment existing feature extraction methods that are based on spectrum analysis. Experimental results show that fractal approaches are effective. When used to augment two traditional features, line and average spectra, fractal approaches led to better classification results. This implies that fractal approaches can capture some information not detected by traditional approaches alone. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1487840]

PACS numbers: 43.60.Lq [JCB]

I. INTRODUCTION

Ship recognition via its radiated sound has attracted attention. Many efforts[1–14] have been made to seek effective features that can lead to good classification of oceanic signals. Due to the complexity of ship sound, it is not easy to obtain the regularities that are essential to ship recognition.

The experimental results achieved here show that fractal approaches are effective for improving ship recognition. These approaches include analysis based on fractional Brownian motion, fractal dimension analysis, and wavelet analysis. In the classification experiments, each approach was independently tested to show how it contributed to classification. Each classification experiment was reported 10 times with randomly selected training and testing samples so that fortuitous conclusions could be excluded. In comparison with the two traditional features, line spectrum and average spectrum, the fractal approaches led to better classification results. This implies that some information that cannot be detected by using traditional approaches can be obtained using fractal approaches.

A major concern about studies based on fractal approaches is whether ship sounds are fractal signals. As practical situations are complex, in a rigorous mathematical sense, ship sounds might not behave exactly as $1/f$ signals. But from an engineering viewpoint, a visible resemblance exists between the spectra of ship sound and those of $1/f$ signals. Through extensive experiments, Urick[15] found that the power spectra of ship sounds decrease 6 decibels per octave at high frequencies. This property is similar to that of $1/f$ signals.[16] For the purpose of classification, it is enough as long as the spectra of ship sounds are similar to those of $1/f$ signals. As shown by the experiments, ship sounds can be effectively classified using fractal approaches. In contrast, active sonar signals cannot be well classified using fractal features because their spectra are distinctly different from those of $1/f$ signals.

The fractal approaches described in this study (1) allow the regularities in ship sound to be viewed in a new way; (2) provide some information not provided by traditional features; and (3) improve ship recognition. Because no single approach can capture all useful information, different features working in collaboration with each other can provide complementary information. As shown in this study, some classes that cannot be well classified using traditional features can be well classified using fractal features.

II. CLASSIFIER

It is assumed that there are M classes $H_1, H_2, \ldots, H_M$, and that each class is composed of $2N$ feature vectors. Let $T_i=\{U_{i,1}, U_{i,2}, \ldots, U_{i,N}\}$ and $S_i=\{V_{i,1}, V_{i,2}, \ldots, V_{i,N}\}$, respectively, denote N training and testing samples that are randomly selected from H_i, where $T_i \cap S_i = \phi$ and $T_i \cup S_i = H_i$. The mean vector of T_i is defined as

$$\mu_i = \frac{1}{N} \sum_{j=1}^{N} U_{i,j}. \tag{1}$$

In the training stage, $\{\mu_i | i=1,2,...,M\}$ should be calculated. In the testing stage, $\{D(V_{i,j}, \mu_k) | k=1,2,...,M;\ j=1,2,...,M;\ j=1,2,...,N\}$ should be computed, where $D(V_{i,j}, \mu_k)$ represents the Mahalonobis distance[17] between

[a)]Electronic mail: yangsu@nju.edu.cn
[b)]Electronic mail: lizshun@nwpu.edu.cn
[c)]Electronic mail: xlwang@nju.edu.cn

 0001-4966/2002/112(1)/172/6/$19.00

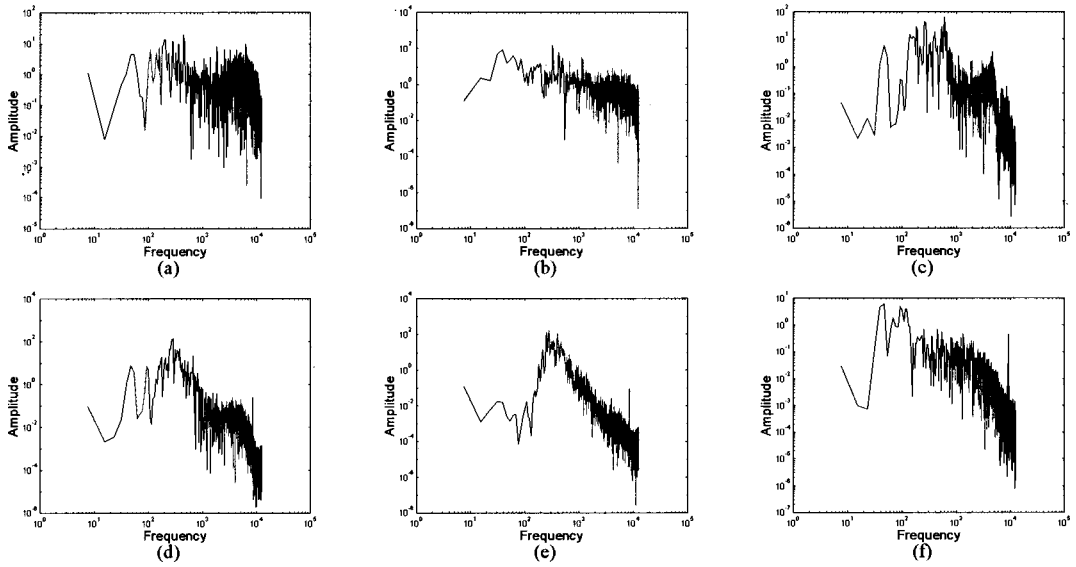


FIG. 1. (a) Power spectrum of ship A sample in log–log plot. (b) Power spectrum of ship B sample in log–log plot. (c) Power spectrum of ship C sample in log–log plot. (d) Power spectrum of ship D sample in log–log plot. (e) Power spectrum of ship E sample in log–log plot. (f) Power spectrum of ship F sample in log–log plot.

V_{ij} and μ_k. If $l=\arg\min_k\{D(V_{i,j},\mu_k)\}$, it can be decided that $V_{i,j}$ belongs to H_l.

III. DATA

The data set used in the experiments contains the sound samples radiated from six ships. With regard to each ship, 100 samples were collected. The data length of each sample is 3264. In order to compensate for the distance variation between ship and receiver, each sample was normalized to possess unit energy by dividing the square root of its energy. The power spectra of the representative samples for each ship are illustrated, respectively, in log–log plots in Fig. 1. Figure 2 illustrates the power spectrum of a fractional Brownian motion (fBm) sample selected from 100 fBm samples that were generated using a Fourier filtering method.[18] Figures 1 and 2 show that the spectra of the ship sound samples are visibly similar to that of the fBm sample.

IV. FRACTIONAL BROWNIAN MOTION BASED ANALYSIS

The power spectra of ship sound, which decrease 6 decibels per octave at high frequencies,[15] are similar to those of fractional Brownian motion (fBm).[19] This motivates us to apply the tools that are useful in analyzing fBm to ship sound analysis. Let $B_H(t)$ denote a fBm with parameter H. The differential sequences of $B_H(t)$ satisfy the T^H law,[19]

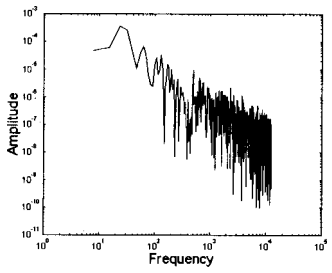


FIG. 2. Power spectrum of fBm sample in log–log plot.

$$\sigma_k=\sqrt{\mathrm{var}[B_H(t+k\Delta)-B_H(t)]}=C_H(k\Delta)^H, \tag{2}$$

where Δ represents the sampling interval, k is the step to form the differential sequence $B_H(t+k\Delta)-B_H(t)$, and C_H can be deemed as a constant if H is given. Let

$$H_k=[\log\sigma_{k+1}-\log\sigma_k]/[\log(k+1)-\log k]. \tag{3}$$

By substituting Eq. (2) into Eq. (3), it follows that $H_k=H$. Different H corresponds to different fBm model. In correspondence with H, $\{H_k\}$ and $\{\sigma_k\}$ are taken as two kinds of feature vectors in this study. In practice, ship sounds could not be ideal fBm signals. Thus, H_k could not be constant at each step k, and σ_k might not satisfy T^H law exactly. Even so, as confirmed by the experiments, $\{H_k\}$ and $\{\sigma_k\}$ provide essential information for ship recognition.

Supposing $s_i:\ i=1,2,...,N$ is a ship sound sequence, the mean value and the standard deviation of its differential sequence $s_{i+k}-s_i:\ i=1,2,...,N-k$ are, respectively,

$$\mu_k=\frac{1}{N-k}\sum_{i=1}^{N-k}(s_{i+k}-s_i), \tag{4}$$

$$\sigma_k=\sqrt{\frac{1}{N-k}\sum_{i=1}^{N-k}(s_{i+k}-s_i-\mu_k)^2}. \tag{5}$$

By substituting Eq. (5) into Eq. (3), the two feature vectors $\{\sigma_k\}$ and $\{H_k\}$ can be calculated. Figures 3 and 4, respectively, show $\{\sigma_k\}$ and $\{H_k\}$ of the sound samples from the six ships. Both the consistency between the samples from each identical ship, and the divisibility between the samples from every two different ships, can be observed in Figs. 3 and 4.

V. FRACTAL DIMENSION ANALYSIS

As a geometrical tool, a fractal dimension can be used to characterize the waveform details of a signal. Among the many definitions of fractal dimension, the ''blanket-covering'' dimension is suitable for feature extraction[20,21] be-

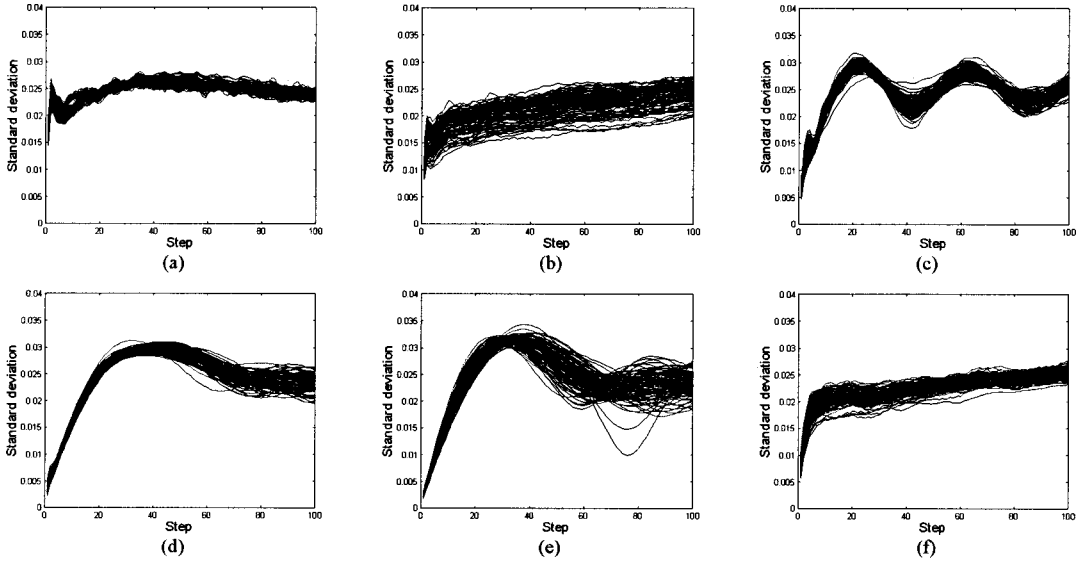


FIG. 3. (a) σ_k of ship A samples, (b) σ_k of ship B samples, (c) σ_k of ship C samples, (d) σ_k of ship D samples, (e) σ_k of ship E samples, (f) σ_k of ship F samples.

cause its computation is invariant to signal shifting in the coordinates of time and amplitude.[22] For a time sequence $f(n)$: $n=0,1,...,N$, choosing a scale r to form an upper envelope $U_r(n)$ and a lower envelope $L_r(n)$, where

$$U_r(n)=\max\{U_{r-1}(n-1),U_{r-1}(n)+1,U_{r-1}(n+1)\}, \tag{6}$$

$$L_r(n)=\min\{L_{r-1}(n-1),L_{r-1}(n)-1,L_{r-1}(n+1)]\}, \tag{7}$$

$$U_0(n)=L_0(n)=f(n), \tag{8}$$

the fractal measurement can be calculated by

$$L(r)=\frac{1}{2r}\sum_{n=0}^{N}[U_r(n)-L_r(n)]. \tag{9}$$

In ideal conditions, $\log L(r)$ and $\log r$ should satisfy the following linear relation:

$$\log L(r)=(1-D)\log r+\log K, \tag{10}$$

where D is the so-called ''blanket-covering'' dimension and K is a constant. Then, D can be acquired by fitting $\log L(r)$ with respect to $\log r$ in a least-square sense. In practice, since $\log L(r)$ versus $\log r$ is not a straight line but a smooth curve, the fractal dimension is not constant at each scale r. Accordingly, the fractal dimension at a given scale r is computed via

$$D_r=1-[\log L(r+1)-\log L(r)]/[\log(r+1)-\log r], \tag{11}$$

which can be derived from Eq. (10). Here $\{D_r\}$ is used as one kind of feature vector in this study. Figure 5 shows $\{D_r\}$ for the samples from six ships. The divisibility between every two different classes, and the consistency within each identical class, visibly exist in Fig. 5. Also, $\{D_r\}$ for the 100

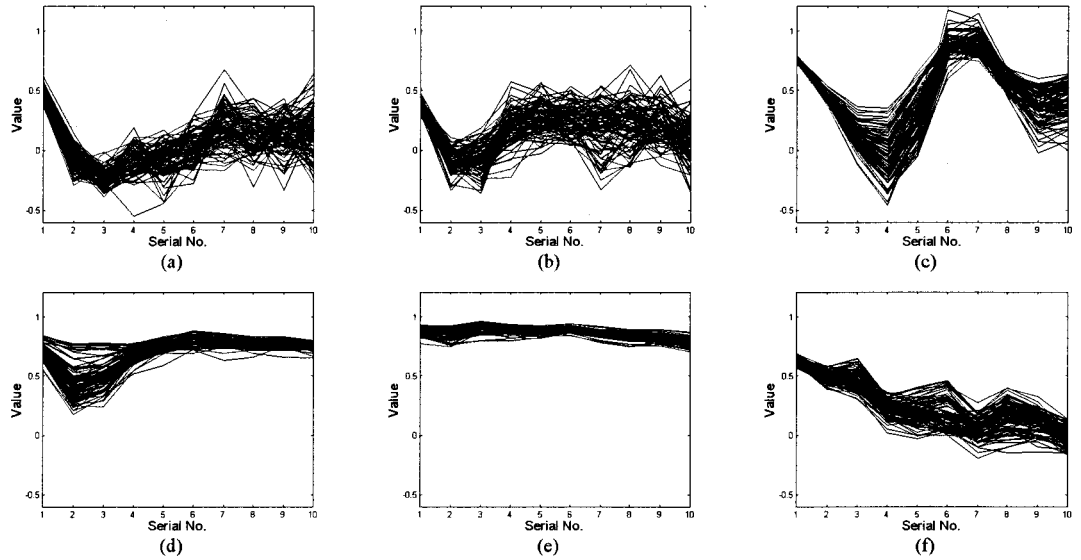


FIG. 4. (a) H_k of ship A samples, (b) H_k of ship B samples, (c) H_k of ship C samples, (d) H_k of ship D samples, (e) H_k of ship E samples, (f) H_k of ship F samples.

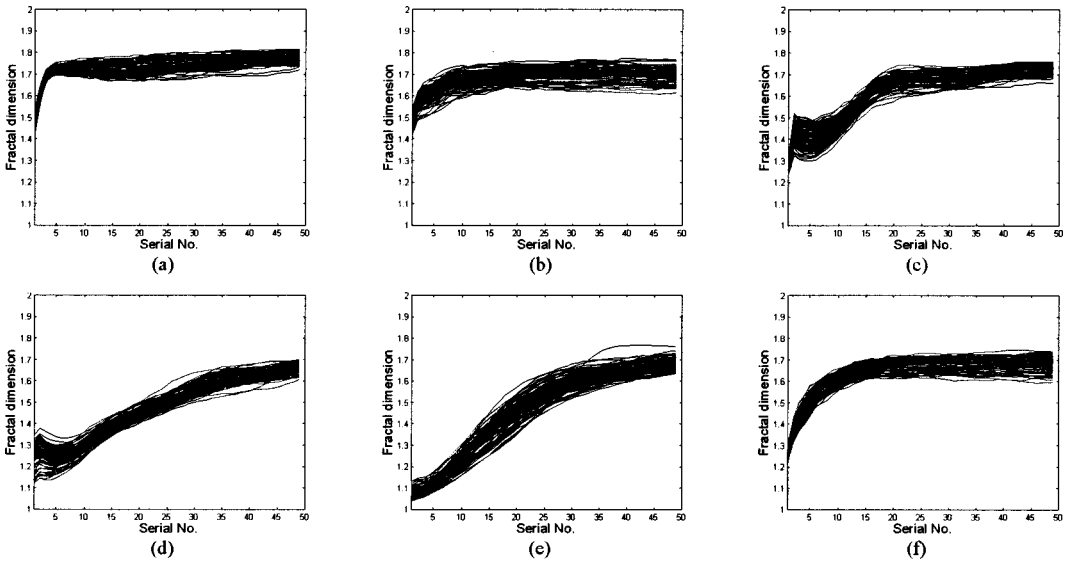


FIG. 5. (a) D_r of ship A samples, (b) D_r of ship B samples, (c) D_r of ship C samples, (d) D_r of ship D samples, (e) D_r of ship E samples, (f) D_r of ship F samples.

fBm samples that were generated using the Fourier filtering method are illustrated in Fig. 6. In both Figs. 5 and 6, D_r versus r ascends at small scales and is nearly constant at big scales. It indicates that $\log L(r)$ versus $\log r$ is not a straight line in practice, even for fBm samples. Therefore, it is reasonable to utilize $\{D_r\}$ for the purpose of classification.

VI. WAVELET ANALYSIS

Wavelet analysis is a useful tool for analyzing $1/f$ signals.[23] As the spectra of ship sounds are similar to $1/f$ signals, wavelet analysis is appropriate. The synthesis and analysis of a signal $n(t)$ using a wavelet transform can be expressed, respectively, as

$$n(t)=\sum_m \sum_j n_j^m \psi_j^m(t), \tag{12}$$

$$n_j^m = 2^{j/2}\int_{-\infty}^{+\infty} n(t)\psi_j^m(t)dt, \tag{13}$$

where $\{\psi_j^m(t)\}$ can represent any orthogonal wavelet basis, and $\{n_j^m\}$ denote the corresponding coefficients at scale m. In the following computations, Daubechies wavelets[24] (4 taps) are used. It has been demonstrated that the wavelet coefficients of any given $1/f$ signal satisfy a zero-mean Gaussian distribution,[23] $n_j^m \sim N(0,\sigma_m^2)$. It means that the distinctions between different $1/f$ models consist in the parameter set $\{\sigma_m^2\}$ for each model. The maximum likelihood estimator of σ_m^2 is[23]

$$\hat{\sigma}_m^2 = \frac{1}{N(m)}\sum_j (n_j^m)^2, \tag{14}$$

where $N(m)$ represents the number of wavelet coefficients at scale m. Because the estimated parameter set $\{\hat{\sigma}_m^2\}$ is the key to identify the signal model that generates a given sample $n(t)$, it is used as one kind of feature vector in this study.

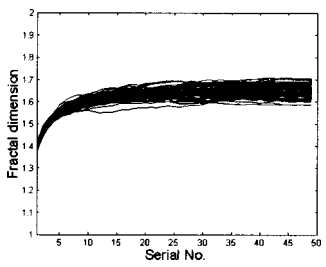


FIG. 6. D_r of fBm samples.

VII. LINE SPECTRUM

The line spectrum is a widely used traditional feature in ship sound recognition. The feature extraction method is as follows. First, the frequencies at which line components frequently occur should be selected through the following procedure.

(1) Compute the power spectrum of each sample.
(2) In the power spectrum of each sample, locate the points that have local minimum amplitudes and denote them as valleys.
(3) Within the region between each pair of adjacent valleys, select the point that has the maximum amplitude and denote such a point as a peak.
(4) Regarding the samples in each class, if the probability that a peak occurs at a given frequency exceeds 60%, this frequency is selected as a line spectral frequency.

Second, for each sample, let the amplitude of the peak at each line spectral frequency be a feature value and take the feature values at all the line spectral frequencies to construct the feature vector of the line spectrum. In view of possible frequency shifting of line components, if there are several peaks within a small range around a line spectral frequency, the maximum amplitude of these peaks is selected as the feature value at this line spectral frequency. If no peak occurs within the small region around a given line spectral frequency, let the feature value at this line spectral frequency be

TABLE I. Classification results using different features.

Feature	Classification accuracy (%)						
	A	B	C	D	E	F	Total
FBM 1	99.6	100	100	98.2	93.2	100	98.5
FBM 2	96.8	100	100	99.8	94	100	98.43
FD	97.4	100	100	99.8	96.6	100	98.97
WA	100	100	100	100	95.6	99.8	99.23
LS	94.4	93.8	89.8	88.6	78.6	72	86.2
AS	100	97.8	94.4	81.6	80.2	100	92.33

0. Here, the half-width of such a small region is allowed to be at most one interval in the frequency axis.

VIII. AVERAGE SPECTRUM

Unlike line components, which are mainly located at low frequencies, the average spectrum reflects the characteristics of a signal over the entire frequency domain. Computation of average spectrum is as follows. First, compute the power spectrum of a given signal using Welch's averaged periodogram method,[25] where a Hanning window is used, and the computation is based on a four-segment average without overlay. Second, divide the whole frequency axis into several equal parts without overlay. Third, let the mean amplitude of each part be a feature value, and take the feature values of all the parts to construct the feature vector of the average spectrum. In the following classification experiments, the number of points included in each part is 50.

IX. CLASSIFICATION EXPERIMENTS

Here, the two fBm feature vectors $\{\sigma_k|k=1,2,...,10\}$ and $\{H_k|k=1,2,...,10\}$, the fractal dimension feature vector $\{D_r|r=1,2,...,10\}$, the wavelet feature vector $\{\hat{\sigma}_m^2\}$, the line spectrum feature vector, and the average spectrum feature vector, which are denoted as FBM1, FBM2, FD, WA, LS, and AS, respectively, are used to classify the data set. Each single feature is used independently to classify the data set. The classification experiment using the same feature is repeated 10 times, in each of which 50 samples are randomly selected from the 100 samples of each ship to construct the training set and the other 50 samples of each ship are used to form the testing set. To compare the performance of each feature, the average classification accuracy of the 10 tests with regard to each ship and the overall classification accuracy regarding the six ships are summarized in Table I.

The performance of LS is obviously inferior to that of any other feature. Besides, AS is apparently inferior to the fractal features in distinguishing ships D and E. That is the major reason why the overall performance of AS is worse than that of any fractal feature in this study. The above comparison shows that fractal features can capture information no detected by traditional features.

From anther viewpoint, σ_k in FBM1 can be understood as the variance of a ship sound passing through a filter whose transfer function is $H(\omega)=(1-e^{-jk\omega})$. Meanwhile, WA concentrates on constant-Q bands while AS focuses on bands with equivalent width. The band partition of FBM1 and WA aims at $1/f$ signals, while that of AS is for more general cases. As the spectra of ship sound are close to those of $1/f$ signals, the performance of FBM1 and WA is thus better than that of AS.

X. SUMMARY

Due to the complexity of ship sound, ship recognition via its radiated sound has been a difficult problem to date. This paper concerns a new approach that can improve ship recognition. The contribution of the fractal based feature extraction methods to ship recognition is confirmed by this study. The essential points in this study are as follows:

(1) Four fractal features, which include two fBm based features, the fractal dimension, and a wavelet feature, are proposed to augment the two traditional features, line and average spectra.
(2) The following regularities can be observed in the experiments. The fractal features of the samples from every identical ship are similar to each other while the distinctions among different ships are obvious.
(3) In view of the classification results, the contribution of fractal features to ship recognition is shown to exist. Since each approach has been tested independently through 10 experiments with randomly selected training and testing samples, the effectiveness of each fractal feature is substantially confirmed.
(4) According to the experimental results, some samples that cannot be well classified by using traditional features can be satisfactorily classified using fractal features. Fractal features thus provide essential information not provided by traditional features.
(5) Fractal approaches allow the characteristics of ship sound to be viewed in a new way compared to traditional methods and then augment existing ways for ship recognition.

ACKNOWLEDGMENT

The authors are very grateful to Professor John C. Burgess, the associate editor, for his beneficial comments in improving the technical presentation and great help in improving the language of this paper. This work is partially supported by the National Science Foundation of China for Distinguished Young Scholars under Grant No. 19925414 and the National Science Foundation of China under Grant No. 19834040.

[1] T. J. Hayward, "Classification by multiple-resolution statistical analysis with application to automated recognition of marine mammals sound," J. Acoust. Soc. Am. **101**, 1516–1526 (1997).
[2] F. B. Shin and D. H. Kil, "Full-spectrum signal processing using a classify-before-detect paradigm," J. Acoust. Soc. Am. **99**, 2188–2197 (1996).
[3] M. L. Barlett, K. W. Baugh, and G. R. Wilson, "Transient detection using the nonstationary bispectrum," J. Acoust. Soc. Am. **99**, 3018–3028 (1996).
[4] P. J. Loughlin and B. Tacer, "On the amplitude- and frequency-modulation decomposition of signals," J. Acoust. Soc. Am. **100**, 1594–1601 (1996).
[5] P. J. Sherman and L. B. White, "Improved periodic spectrum analysis with application to diesel vibration data," J. Acoust. Soc. Am. **98**, 3285–3301 (1995).
[6] A. M. Richardson and W. S. Hodgkjiss, "Bispectruml analysis of under-

water acoustic data,'' J. Acoust. Soc. Am. **96**, 828–837 (1994).
[7] J. R. Potter, D. K. Mellinger, and C. W. Clark, ''Marine mammal call discrimination using artificial neural networks,'' J. Acoust. Soc. Am. **96**, 1255–1262 (1994).
[8] J. Meister, ''A neural network harmonic family classifier,'' J. Acoust. Soc. Am. **93**, 1488–1495 (1993).
[9] L. A. Pflug, G. E. Loup, J. W. Loup, and R. L. Fiell, ''Properties of high-order correlations and spectrum for bandlimited deterministic transients,'' J. Acoust. Soc. Am. **91**, 975–988 (1992).
[10] M. J. Hinich, D. Marandino, and E. J. Sullivan, ''Bispectrum of ship-radiated noise,'' J. Acoust. Soc. Am. **85**, 1512–1517 (1989).
[11] T. L. Hemminger and Yoh-Han Pao, ''Detection and classification of underwater acoustic transients using neural networks,'' IEEE Trans. Neural Netw. **5**, 712–717 (1994).
[12] A. Kundu, G. C. Chen, and C. E. Persons, ''Transient sonar signal classification using hidden Markov models and neural nets,'' IEEE J. Ocean. Eng. **19**, 87–99 (1994).
[13] J. Ghosh, L. M. Deuser, and S. D. Beck, ''A neural network based hybrid system for detection, characterization, and classification of short-duration oceanic signals,'' IEEE J. Ocean. Eng. **17**, 351–363 (1992).
[14] I. P. Kirstcins, S. K. Mchta, and J. Pay, ''Adaptive separation of unknown narrowband and broadband time series,'' IEEE ICASSP, Seattle **4**, 2525–2528 (1998).
[15] R. J. Urick, *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York, 1983), pp. 328–353.
[16] M. S. Keshner, ''1/*f* noise,'' Proc. IEEE **70**, 212–218 (1982).
[17] B. D. Ripley, *Pattern Recognition and Neural Networks* (Cambridge University Press, Cambridge, 1996).
[18] H. O. Peitgen and D. Saupe, *The Science of Fractal Images* (Springer-Verlag, New York, 1988).
[19] B. B. Mandelbrot and H. W. Van Ness, ''Fractional Brownian motions, fractional noises and applications,'' SIAM Rev. **10**, 422–436 (1968).
[20] D. R. Carmichael, L. M. Linnett, S. J. Clarke, and B. R. Calder, ''Seabed classification through multifractal analysis of sidescan sonar imagery,'' IEE Proc. Radar, Sonar Navig. **143**, 140–148 (1996).
[21] S. Yang, Z. S. Li, and X. L. Wang, ''Vessel radiated noise recognition with fractal features,'' Electron. Lett. **36**, 923–924 (2000).
[22] P. Maragos and Fang-Kuo Sun, ''Measuring the fractal dimension of signals: Morphological covers and iterative optimization,'' IEEE Trans. Signal Process. **41**, 108–121 (1993).
[23] G. Wornell, *Signal Processing with Fractals: A Wavelet-Based Approach* (Prentice-Hall, Upper Saddle River, NJ, 1996), pp. 1–111.
[24] I. Daubechies, *Ten Lectures on Wavelets* (SIAM, Philadelphia, 1992).
[25] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1975).

Combined Helmholtz equation–least squares method for reconstructing acoustic radiation from arbitrarily shaped objects

Sean F. Wu and Xiang Zhao
Department of Mechanical Engineering, Wayne State University, Detroit, Michigan 48202

(Received 27 August 2001; revised 19 April 2002; accepted 25 April 2002)

A combined Helmholtz equation–least squares (CHELS) method is developed for reconstructing acoustic radiation from an arbitrary object. This method combines the advantages of both the HELS method and the Helmholtz integral theory based near-field acoustic holography (NAH). As such it allows for reconstruction of the acoustic field radiated from an arbitrary object with relatively few measurements, thus significantly enhancing the reconstruction efficiency. The first step in the CHELS method is to establish the HELS formulations based on a finite number of acoustic pressure measurements taken on or beyond a hypothetical spherical surface that encloses the object under consideration. Next enough field acoustic pressures are generated using the HELS formulations and taken as the input to the Helmholtz integral formulations implemented through the boundary element method (BEM). The acoustic pressure and normal component of the velocity at the discretized nodes on the surface are then determined by solving two matrix equations using singular value decomposition (SVD) and regularization techniques. Also presented are in-depth analyses of the advantages and limitations of the CHELS method. Examples of reconstructing acoustic radiation from separable and nonseparable surfaces are demonstrated. [DOI: 10.1121/1.1487845]

PACS numbers: 43.60.Pt, 43.40.Rj, 43.50.Yw [EGW]

I. INTRODUCTION

The Helmholtz equation least squares (HELS) method has shown a great promise in reconstructing acoustic fields in both exterior[1–4] and interior[5] regions. The main advantages of the HELS method are its simplicity, flexibility, and efficiency. It expresses the field acoustic pressure as an expansion of basis functions that are the particular solutions to the Helmholtz equation. The coefficients associated with these basis functions are determined by matching the assumed-form solution to the measured acoustic pressures in the field. The errors incurred in this process are minimized by the least squares and the solutions thus obtained are unique. Since there are no restrictions on the measurement and reconstruction locations, implementation of the HELS method is very easy and flexible. This method has been used successfully to reconstruct acoustic radiation from a simplified engine block[2] and a vehicle front end.[3] For a spherical surface, it is especially effective: it enables one to visualize the entire acoustic field based on a finite number of acoustic pressure measurements taken over a planar surface on one side.[4]

However, the accuracy of reconstruction on the surface of an arbitrarily shaped object may be unsatisfactory. This is because the expansion using the spherical outgoing waves for an acoustic field on an arbitrarily shaped surface is incomplete. The more the surface deviates from a sphere, the worse the reconstruction accuracy becomes. While the HELS solutions can be exact outside a minimum imaginary sphere that encloses an arbitrary object, they are approximate inside this sphere. The further the reconstruction points go inward, the worse the HELS solutions become. Hence the HELS method is in general not suitable for reconstructing acoustic quantities on the surface of a highly non-spherical object.

An alternative for reconstructing acoustic radiation from an arbitrary object is the Helmholtz integral theory.[6] In implementing this integral theory, BEM is used and the surface is discretized into segments and the acoustic field is specified on the nodes of these segments using a particular interpolation scheme. This BEM-based NAH has been applied successfully to reconstructing acoustic radiation from structures in both exterior[7–10] and interior[11–13] regions.

The main advantage of the BEM-based NAH is its generality, allowing one to tackle an arbitrarily shaped object. The disadvantage is that the surface Helmholtz integral equation may fail to yield unique solutions for the exterior problems when the excitation frequencies are close to the eigenfrequencies of the corresponding boundary value problems in the interior region. While this non-uniqueness difficulty can be overcome by the CHIEF method,[14] the efficiency and accuracy of reconstruction may be greatly affected.

The main drawback of the BEM-based NAH, however, is due to the fact that the acoustic field is reconstructed through a spatial sampling. In other words, one must have a minimum number of nodes per wavelength in order to achieve a desired resolution in reconstruction. Accordingly, one must take enough measurements of field acoustic pressures to determine the acoustic quantities defined on these nodes. For a complex structure vibrating at low-to-mid frequencies, the number of nodes necessary to describe the surface acoustic field can be large. Hence the number of measurements is large, which makes the reconstruction process very time consuming. While there are techniques developed recently to avoid the singularity problem inherent in the Helmholtz integral equation[15] and methodologies to optimize the measurement locations using an effective independence,[16] implementation of the BEM-based NAH is

not straightforward and reconstruction of the acoustic field is often slow and inefficient.

It is emphasized that in using the BEM-based NAH it is not necessary to have exactly the same measurements as discrete nodes. This is because with SVD and regularization procedures, one can have either over- or under-determined systems of equations. Accordingly, one can use fewer measurements than the discrete nodes. However, the accuracy of reconstruction cannot be guaranteed if the measurements are substantially fewer than the discrete nodes. This is because the measured data are not error free and background noises are always present. In order to obtain a convergent solution, the matrix must be truncated to filter out the evanescent waves that fall under the background noises. If measurements are too few, the equivalent cutoff wave number is forced to be set very low. As a result, the high spatial wavenumber contents are filtered out and aliasing occurs in reconstruction. Such an example is illustrated in Sec. IV.

In this paper it is shown that by combining the HELS method and BEM-based NAH, known as the CHELS method, the efficiency of reconstruction can be significantly enhance and satisfactory reconstruction be obtained with relatively few measurements. The limitations of the CHELS method are also discussed. Numerical examples of reconstruction of acoustic radiation from separable and nonseparable geometry are demonstrated. In particular, the surface acoustic pressures reconstructed by the CHELS method are compared with those obtained by the BEM-based NAH. Results show that the CHELS method can yield satisfactory reconstruction with much fewer measurements than the BEM-based NAH does. When the same few measurements are used in the BEM-based NAH, aliasing may occur and the reconstructed surface acoustic pressures may be distorted.

II. RECONSTRUCTION FORMULATIONS

Consider sound radiation from an arbitrary object in an unbounded fluid medium of ambient density ρ_0 and speed of sound c. Using the HELS method, one can write the field acoustic pressure $p(\vec{x},\omega)$ at any point $\vec{x}$ as[2]

$$p(\vec{x},\omega)=\sum_{j=1}^{J} C_j \Psi_j(\vec{x},\omega), \tag{1}$$

where Ψ_j are the basis functions that can be expressed in the spheroid coordinate systems. For example, using the spherical coordinates we can write Ψ_j as[17]

$$\Psi_j=\psi_{n,l}(r,\theta,\phi)=h_n(kr)P_{n,l}(\cos\theta)e^{il\phi}, \tag{2}$$

where $k=\omega/c$ is the acoustic wave number, h_n are the spherical Hankel functions, and $P_{n,l}$ are the associated Legendre functions, here the indices j, n, and l are related by $j=n^2+n+l+1$ with l varying from $-n$ to $+n$. The coefficients C_j can be determined by matching the assumed-form solution (1) to the measured acoustic pressures $p(\vec{x}_m,\omega)$ at $\vec{x}_m$, $m=1,2,...,M$ $(M\geqslant J)$. In a matrix form, we can write this process as

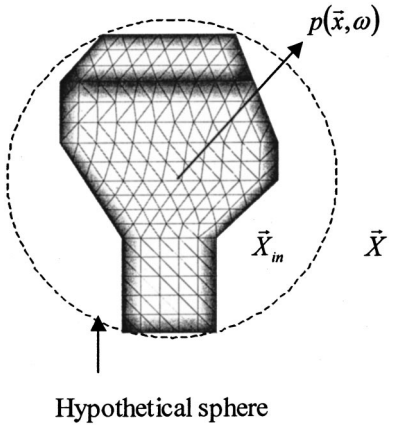


FIG. 1. Schematic of reconstruction of acoustic radiation from an arbitrary object.

$$\{p(\vec{x}_m,\omega)\}_{M\times 1}=[\Psi]_{M\times J}\{C\}_{J\times 1}. \tag{3}$$

Equation (3) can be solved by singular value decomposition (SVD) if the matrix is ill conditioned. Otherwise, the coefficients C_j can be obtained using a pseudoinversion

$$\{C\}_{J\times 1}=([\Psi]^{\mathrm{T}}_{M\times J}[\Psi]_{M\times J})^{-1}[\Psi]^{\mathrm{T}}_{M\times J}\{p(\vec{x}_m,\omega)\}_{M\times 1}. \tag{4}$$

Once the coefficients C_j are specified, the acoustic pressure can be reconstructed using Eq. (1). Because the basis functions Ψ_j defined by Eq. (2) describe the outgoing spherical waves, solutions thus obtained are accurate for $\vec{x}\in\vec{X}$ outside a hypothetical spherical surface that encloses the object under consideration (see Fig. 1). In fact, if the measured acoustic pressures are exact, then the reconstructed pressure $p(\vec{x},\omega)$ converges to the true value as the number of expansion $J\to\infty$.[18]

When the reconstruction point is inside this minimum sphere $\vec{x}\in\vec{X}_{in}$, however, the expansion of the basis functions Ψ_j is incomplete and the solutions of Eq. (1) are approximate.[19] In other words, the acoustic field inside the minimum sphere cannot be described by the outgoing spherical waves alone. The more $\vec{x}$ moves inward, the worse the accuracy of reconstruction is.

Examples of such are demonstrated by Rayess and Wu,[20] in which the acoustic fields radiated from the prolate and oblate spheroids are expressed as superposition of the outgoing spherical waves. Results show that outside the minimum spheres that enclose the prolate and oblate spheroids, the acoustic fields can be reconstructed exactly using the spherical Hankel functions and spherical harmonics. Inside the minimum sphere, the accuracy of reconstruction is found to depend on the aspect ratios of the prolate and oblate spheroids. The larger the aspect ratio is, the worse the accuracy of reconstruction becomes. While the real parts of the spherical Hankel functions and spherical harmonics agree exactly with those of the analytic solutions for the prolate and oblate spheroids, the imaginary parts of these functions may grow unboundedly as the field point $\vec{x}$ approaches certain locations of the prolate and oblate spheroids.[20] This is

why the accuracy of reconstruction of the acoustic field on the surface of an arbitrary object using Eq. (1) may be unsatisfactory.

A general method for reconstructing acoustic radiation from an arbitrarily shaped object is via the Helmholtz integral theory, which correlates the field acoustic pressure $p(\vec{x},\omega)$ to the surface acoustic quantities

$$p(\vec{x},\omega)=\frac{1}{4\pi}\int_S\int\left[p(\vec{x}_S,\omega)\frac{\partial G(\vec{x}|\vec{x}_S)}{\partial n}-G(\vec{x}|\vec{x}_S)\frac{\partial p(\vec{x}_S,\omega)}{\partial n}\right]dS, \tag{5}$$

where $G=e^{ikR}/R$ is the free-space Green's function with $R=|\vec{x}-\vec{x}_S|$ being the distance between the field and surface points, $\partial G/\partial n$ represents a normal derivative of G with respect to the surface unit normal. The normal derivative of the surface acoustic pressure is related to the normal component of the surface velocity v_n via the Euler's equation, $\partial p(\vec{x}_S,\omega)/\partial n=i\omega\rho_0 v_n(\vec{x}_S,\omega)$.

The surface acoustic pressure $p(\vec{x}_S,\omega)$ and the normal component of the surface velocity $v_n(\vec{x}_S,\omega)$ are related through the Helmhltz integral equation that can be obtained by taking the limit as the field point approaches the surface $\vec{x}\to\vec{x}_{S'}$,

$$p(\vec{x}_{S'},\omega)=\frac{1}{2\pi}\int_S\int\left[p(\vec{x}_S,\omega)\frac{\partial G(\vec{x}_{S'}|\vec{x}_S)}{\partial n}-G(\vec{x}_{S'}|\vec{x}_S)\frac{\partial p(\vec{x}_S,\omega)}{\partial n}\right]dS. \tag{6}$$

Equations (5) and (6) correlate the field acoustic pressure to surface acoustic quantities. For an arbitrary object there are no closed-form solutions to Eqs. (5) and (6), hence they are solved numerically. Suppose that the surface is discretized into segments with N nodes, then we can rewrite Eqs. (5) and (6) as

$$p(\vec{x},\omega)=\{T_p\}_{1\times N}\{p(\vec{x}_S,\omega)\}_{N\times 1}, \tag{7a}$$

$$p(\vec{x},\omega)=\{T_v\}_{1\times N}\{v_n(\vec{x}_S,\omega)\}_{N\times 1}, \tag{7b}$$

where $\{T_p\}_{1\times N}$ and $\{T_v\}_{1\times N}$ relate the field acoustic pressure to the surface acoustic pressure and the normal component of the surface velocity, respectively, which are defined by

$$\{T_p\}_{1\times N}=(4\pi)^{-1}(\{D\}_{1\times N}+\{M\}_{1\times N}[M_S]^{-1}_{N\times N}\times(2\pi[I]_{N\times N}-[D_S]_{N\times N})), \tag{8a}$$

$$\{T_v\}_{1\times N}=(4\pi)^{-1}(\{D\}_{1\times N}(2\pi[I]_{N\times N}-[D_S]_{N\times N})^{-1}\times[M_S]_{N\times N}+\{M\}_{1\times N}), \tag{8b}$$

where $[I]_{N\times N}$ indicates a unitary matrix, $[M_S]_{N\times N}$ and $[D_S]_{N\times N}$ represent the effects of monopole and dipole on surface points, respectively, and $\{M\}_{1\times N}$ and $\{D\}_{1\times N}$ imply the effects of monopole and dipole on field points, respectively. To determine the surface acoustic pressure and surface normal velocity, we take N measurements $p(\vec{x}_m,\omega)$ to form N simultaneous equations.

$$\{p(\vec{x}_m,\omega)\}_{N\times 1}=[T_p]_{N\times N}\{p(\vec{x}_S,\omega)\}_{N\times 1}, \tag{9a}$$

$$\{p(\vec{x}_m,\omega)\}_{N\times 1}=[T_v]_{N\times N}\{v_n(\vec{x}_S,\omega)\}_{N\times 1}. \tag{9b}$$

Equation (9) enables one to reconstruct surface acoustic pressure and normal component of the surface velocity via an inversion of the matrices $[T_p]_{N\times N}$ and $[T_v]_{N\times N}$. In practice, the measured acoustic pressures $p(\vec{x}_m,\omega)$ are not error free because of the presence of background noises. As a result, the matrix equation (9) may be ill conditioned. To overcome this difficulty, SVD and regularization techniques[21] are used to eliminate the evanescent waves that fall below the background noise level and $p(\vec{x}_S,\omega)$ and $v_n(\vec{x}_S,\omega)$ can be written as

$$\{p(\vec{x}_S,\omega)\}_{N\times 1}=[V]_{N\times N}[\Sigma_p^{-1}]_{N\times N}[U]^{\mathrm{T}}_{N\times N}\{p(\vec{x}_m,\omega)\}_{N\times 1}, \tag{10a}$$

$$\{v_n(\vec{x}_S,\omega)\}_{N\times 1}=[V]_{N\times N}[\Sigma_v^{-1}]_{N\times N}[U]^{\mathrm{T}}_{N\times N}\{p(\vec{x}_m,\omega)\}_{N\times 1}, \tag{10b}$$

where $[V]_{N\times N}[V]^{\mathrm{T}}_{N\times N}=[I]_{N\times N}$ and $[U]_{N\times N}[U]^{\mathrm{T}}_{N\times N}=[I]_{N\times N}$, Σ_p^{-1} and Σ_v^{-1} represent the diagonal matrices that contain inversions of the nonzero singularities of the matrices $[T_p]_{N\times N}$ and $[T_v]_{N\times N}$, respectively.

Equation (10) gives the reconstructed acoustic quantities on the surface of any arbitrary structure. Since these acoustic quantities are described by a spatial discretization, we must have a minimum number of nodes per wavelength in order to avoid aliasing. For a complex structure vibrating at low-to-mid frequencies, the number of discrete nodes needed to describe the surface acoustic quantities can be large. As a result, the number of measurements required to reconstruct these acoustic quantities is large, which makes reconstruction process very time consuming.

III. THE CHELS METHOD

To improve the accuracy of the HELS method and enhance the efficiency of the BEM-based NAH for an arbitrary object, we propose to combine these two methods and describe its procedures as follows.

First, we take M acoustic pressure measurements on or outside a hypothetical spherical surface that encloses the arbitrary object to establish the HELS formulation.

Note that a regularization technique is employed in this process and the optimal number of expansion functions is determined. This is accomplished by dividing the measurements into two groups M_1 and M_2 ($M=M_1+M_2$), where M_1 are taken as the input to the right side of Eq. (4) for solving the coefficients C_j and M_2 used for optimization. Since the least squares method is used, the expansion solution with $J=M_1$ is always the best fit for the M_1 measured quantities. However, it may not yield the best approxima-

tions for the remaining M_2 values simultaneously. This is especially true when the measured acoustic pressures contain errors. The accuracy of reconstruction at any point other than the M_1 locations increases with the number of expansion terms J first, and then deteriorates thereafter.[2]

Hence, to ensure a consistent and satisfactory accuracy of reconstruction, we must find an optimal expansion term J_{op}, which is equivalent to finding a low-pass filter for the spherical harmonics such that the evanescent waves below the background noise level are eliminated. There are many regularization methods available for solving a set of linear equations.[22] The ones that have been tested for the HELS method include an iteration scheme[2] and a constrained minimization[23] and both of them yield approximately the same value of J_{op}.

In this paper, the iteration technique is employed. Namely, M_1 measurements are used to solve for the coefficients C_j, $j=1,2,...,J$, with $J=1$ first. Then Eq. (1) is used to reconstruct the acoustic pressures at all M measurement locations. The sum of the squares of errors of the reconstructed acoustic pressures with respect to the measured data $p(\vec{x}_m,\omega)$, $m=1,2,...,M$, are calculated. Next, the value of J is increased by one and the process is repeated to yield a different sum of the squares of errors. This iteration continues until J_{op} is found that minimizes the sum of the squares of errors with respect to all input data. The value of J_{op} thus obtained is said to be optimized with respect to the given set of measurements and may change with the frequency.

Once J_{op} is specified, the field acoustic pressures is regenerated at as many points as necessary on the hypothetical spherical surface using

$$p(\vec{x},\omega)=\{\Psi\}_{1\times J_{\mathrm{op}}}([\Psi]^{\mathrm{T}}_{M_1\times J_{\mathrm{op}}}[\Psi]_{M_1\times J_{\mathrm{op}}})^{-1}\times[\Psi]^{\mathrm{T}}_{M_1\times J_{\mathrm{op}}}\{p(\vec{x}_m,\omega)\}_{M_1\times 1}. \tag{11}$$

These field acoustic pressures are taken as input to Eq. (10) to reconstruct the surface acoustic pressures and normal component of surface velocities

$$\{p(\vec{x}_S,\omega)\}_{N\times 1}=[V]_{N\times N}[\Sigma_p^{-1}]_{N\times N}[U]^{\mathrm{T}}_{N\times N}[\Psi]_{N\times J_{\mathrm{op}}}\times([\Psi]^{\mathrm{T}}_{M_1\times J_{\mathrm{op}}}[\Psi]_{M_1\times J_{\mathrm{op}}})^{-1}\times[\Psi]^{\mathrm{T}}_{M_1\times J_{\mathrm{op}}}\{p(\vec{x}_m,\omega)\}_{M_1\times 1}, \tag{12a}$$

$$\{v_n(\vec{x}_S,\omega)\}_{N\times 1}=[V]_{N\times N}[\Sigma_v^{-1}]_{N\times N}[U]^{\mathrm{T}}_{N\times N}[\Psi]_{N\times J_{\mathrm{op}}}\times([\Psi]^{\mathrm{T}}_{M_1\times J_{\mathrm{op}}}[\Psi]_{M_1\times J_{\mathrm{op}}})^{-1}\times[\Psi]^{\mathrm{T}}_{M_1\times J_{\mathrm{op}}}\{p(\vec{x}_m,\omega)\}_{M_1\times 1}. \tag{12b}$$

The enhancement in the reconstruction efficiency is obvious. Equation (12) shows that $p(\vec{x}_S,\omega)$ and $v_n(\vec{x}_S,\omega)$ on all N nodes of an arbitrarily shaped surface can now be reconstructed using M_1 measurements. Since $M\ll N$, the measurement time is significantly reduced.

Note that the accuracy of the regenerated field acoustic pressure is consistent with that of the measured data in $\vec{x}_m\in\vec{X}$. This is because the acoustic field in $\vec{x}_m\in\vec{X}$ can be completely and uniquely described by Eq. (1) as $J\to\infty$. The omission of higher order terms, namely, the evanescent waves have a negligibly small impact on the resultant field acoustic pressure. Hence there is no need to take more measurements than necessary. In fact, the accuracy of reconstruction would remain unchanged even if the regenerated field acoustic pressures were replaced by the real measurements. Examples of such are given in the next section for both separable and nonseparable geometry. The trade off is that the accuracy of the reconstructed surface acoustic quantities may be limited because certain evanescent waves may be lost as measurements are taken over a spherical surface rather than a conformal surface at close range.

It is emphasized that one cannot extend the processes discussed above to the interior region of the minimum sphere, either by taking measurements or regenerating the field acoustic pressures in $\vec{x}_m\in\vec{X}_{in}$. This is because the acoustic field in this region cannot be represented adequately by the outgoing spherical waves alone. Numerical experiments have confirmed that while the expansion solution (11) may give the best fit for any set of measurements taken inside the minimum sphere, the accuracy of the regenerated acoustic pressures at other locations may be unsatisfactory and the resulting reconstruction may be distorted.

IV. ANALYSIS OF RECONSTRUCTION RESULTS

In this section, we examine the performance of the CHELS method and compare it with that of the BEM based. In particular, we want to see if the CHELS method can yield satisfactory reconstruction of acoustic radiation accurately and efficiently and to examine its limitations.

A. Separable geometry

The first case studied is a partially vibrating sphere. The reason for selecting this example is because it contains rich evanescent waves and yet the analytic solutions are readily available. The normal surface velocity distribution v_n can be written as

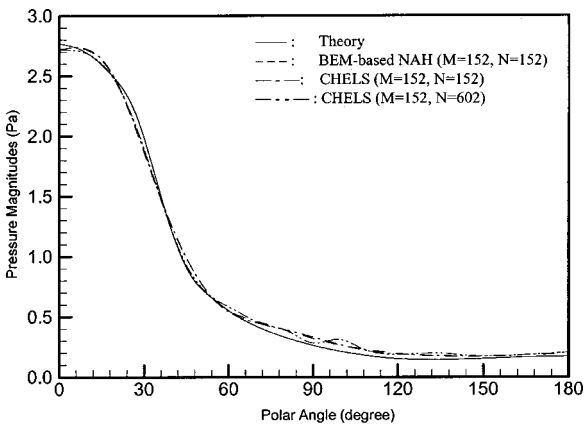


FIG. 2. Comparison of the reconstructed acoustic pressure distributions on the surface of a partially vibrating sphere at $ka=1.46$ based on measurements taken at $r=0.105$ m.

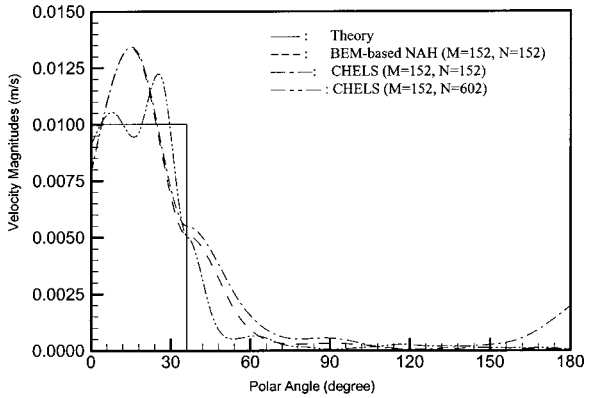


FIG. 3. Comparison of reconstructed normal component of velocity distributions on the surface of a partially vibrating sphere at $ka=1.46$ based on measurements taken at $r=0.105$ m.

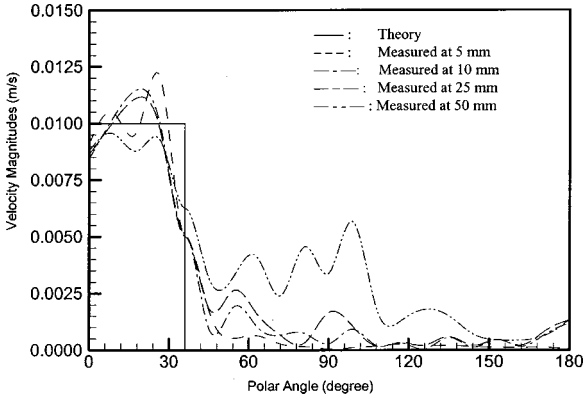


FIG. 5. Comparison of the reconstructed normal component of velocity on the surface of a partially vibrating sphere $ka=1.46$ using the CHELS method based on measurements taken at different radial distances.

$$v_n(a,\theta,\phi)=\begin{cases} v_0, & 0\leqslant|\theta|\leqslant\theta_0, \\ 0, & \text{otherwise}, \end{cases} \tag{13}$$

where v_0 is a constant and the half vertex angle $\theta_0=36°$.

Since the object is a sphere, the minimum surface is a conformal surface. For generality, however, no symmetry property is taken in numerical computations. Assume that the highest dimensionless frequency of interest $(ka)_{\max}=10$, where $a=0.1$ m is the radius of the sphere, then the initial estimate of the number of measurements M is given by[4]

$$M=1.4[(ka)_{\max}+1]^2 \\ =170 \quad \text{for general acoustic radiation.} \tag{14}$$

For convenience, we set $M=152$ so the measurement points can be uniformly distributed over a spherical surface of radius $r=0.105$ m. These measurements are obtained using the formulation given by Morse and Ingard.[24] Half of these data $M_1=76$ are taken as input to Eq. (4) to determine the coefficients C_j and the rest are used to optimize the number of expansion functions. In this case $J_{\text{op}}=26$ is found to be a good choice for $0<ka\leqslant 10$. The field acoustic pressures are then regenerated using Eq. (11).

Note that for a given set of measurements in engineering practice, a larger value of J_{op} implies more evanescent waves captured and therefore, the accuracy of reconstruction may be higher. A smaller value of J_{op} often implies a lower signal to noise ratio and fewer evanescent waves captured. As a result, the reconstructed acoustic field may be unsatisfactory.

As a comparison, we use BEM-based NAH to reconstruct the surface acoustic quantities. To ensure the accuracy of reconstruction, we use six divisions per wavelength to describe the surface acoustic field. Since the wavelength is $\lambda=2\pi a/(ka)$, the average spacing between the neighboring nodes is $\Delta=\pi a/3(ka)$. So at $(ka)_{\max}=10$ the shortest wavelength is approximately $\lambda_{\min}\approx 0.063$ m and $\Delta_{\min}=0.01$ m.

For brevity, we present the results of $ka=1.46$ with an average distance between two neighboring nodes $\Delta=0.032$ m. The corresponding number of nodes using triangular elements and a first-order interpolation is $N=152$. For comparison purpose, we also consider a refined mesh with $N=602$ nodes and an average distance between neighboring nodes of $\Delta=0.016$ m. The acoustic quantities at these nodes are reconstructed by taking the same number of acoustic pressure measurements as that of nodes. These measured data are obtained using the formulation given by Morse and Ingard,[24] which are taken as input to Eq. (10) to reconstruct the surface acoustic pressure and normal component of the surface velocity.

Figure 2 shows the comparison of the reconstructed acoustic pressures at $ka=1.46$ along the generator of the sphere. Results demonstrate that the surface acoustic pressures reconstructed by the CHELS method with $N=152$ and $N=602$ input agree very well with analytic solutions and those given by the BEM-based NAH with 602 input. In particular, it is seen that a fine mesh does not make much difference because the acoustic pressure is continuous.

Refining the mesh sizes does make a significant difference on reconstructing the normal surface velocity distribution however. Figure 3 depicts that a coarse mesh with $N=152$ nodes only enables one to capture the main characteristic of the normal surface velocity distribution. With a fine

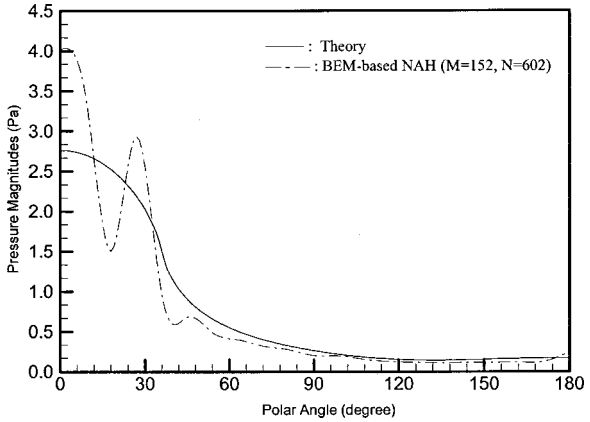


FIG. 4. Reconstructed acoustic pressure distribution on the surface of a partially vibrating sphere at $ka=1.46$ using the BEM-based NAH with 602 discrete nodes and 152 measurements taken at $r=0.105$ m.

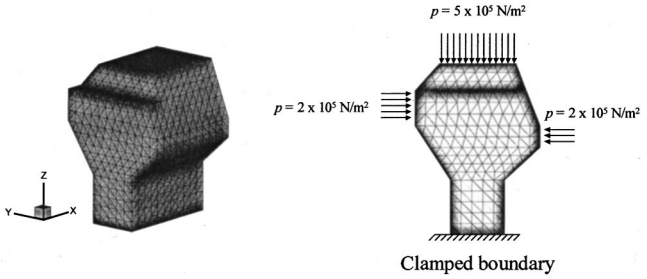


FIG. 6. Schematic of an engine block subject to distributed harmonic force excitations.

mesh of $N=602$ nodes and the same number of input data regenerated by Eq. (11), we can greatly improve the reconstruction accuracy. This is because the surface normal velocity contains more high wavenumber contents than the surface acoustic pressure does.

The fact that the CHELS method can yield satisfactory reconstruction with relatively few measurements is of great significance. It indicates that the fidelity of the input data regenerated by Eq. (11) is preserved. Hence, one does not need to take more measurements than necessary. Moreover, it shows that one can improve the accuracy by increasing the input data. Since these data are calculated but not measured, the efficiency of reconstruction is greatly enhanced.

However, one should not expect the normal surface velocity to converge to the true value even as the number of regenerated input data approaches infinity. This is because the accuracy of reconstruction is controlled by the amount of evanescent waves captured in the measured data. The closer the measurements are to the source surface, the more the evanescent waves are captured, and the more accurate the reconstruction becomes. Once the measurement distance is fixed, so is the amount of the evanescent waves that can be captured. Hence the improvement in the reconstruction accuracy via increasing the number of regenerated input data is limited.

Also it is emphasized that one does not need to have exactly the same measurements as the discrete nodes. This is because by using SVD and regularization, one can have an over- or under-determined system of equations, or equivalently, have more or fewer measurements than the discrete

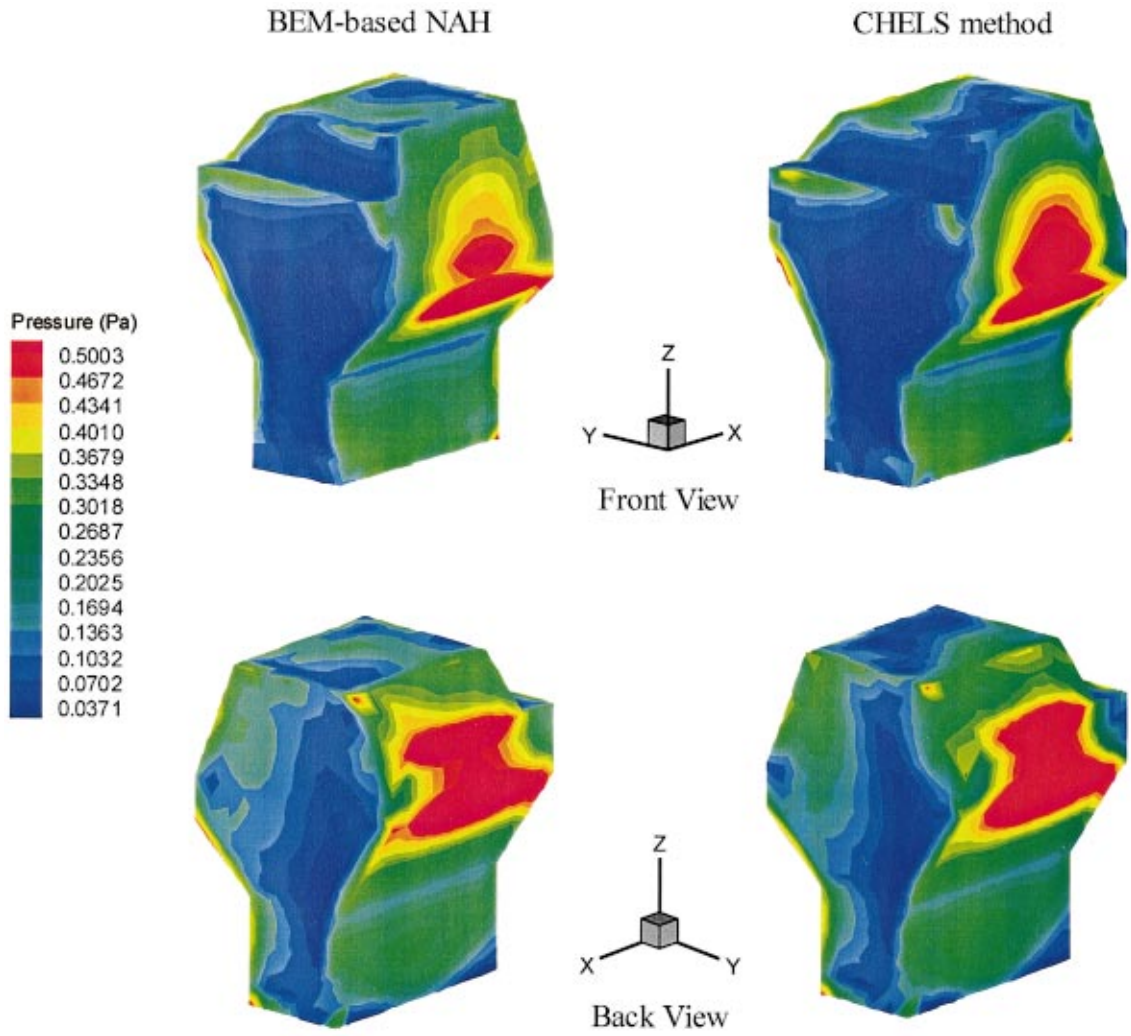


FIG. 7. Comparison of reconstructed acoustic pressure distributions over the entire surface of the engine block using the CHELS method (right column) and BEM-based NAH (left column) at $ka=1$.

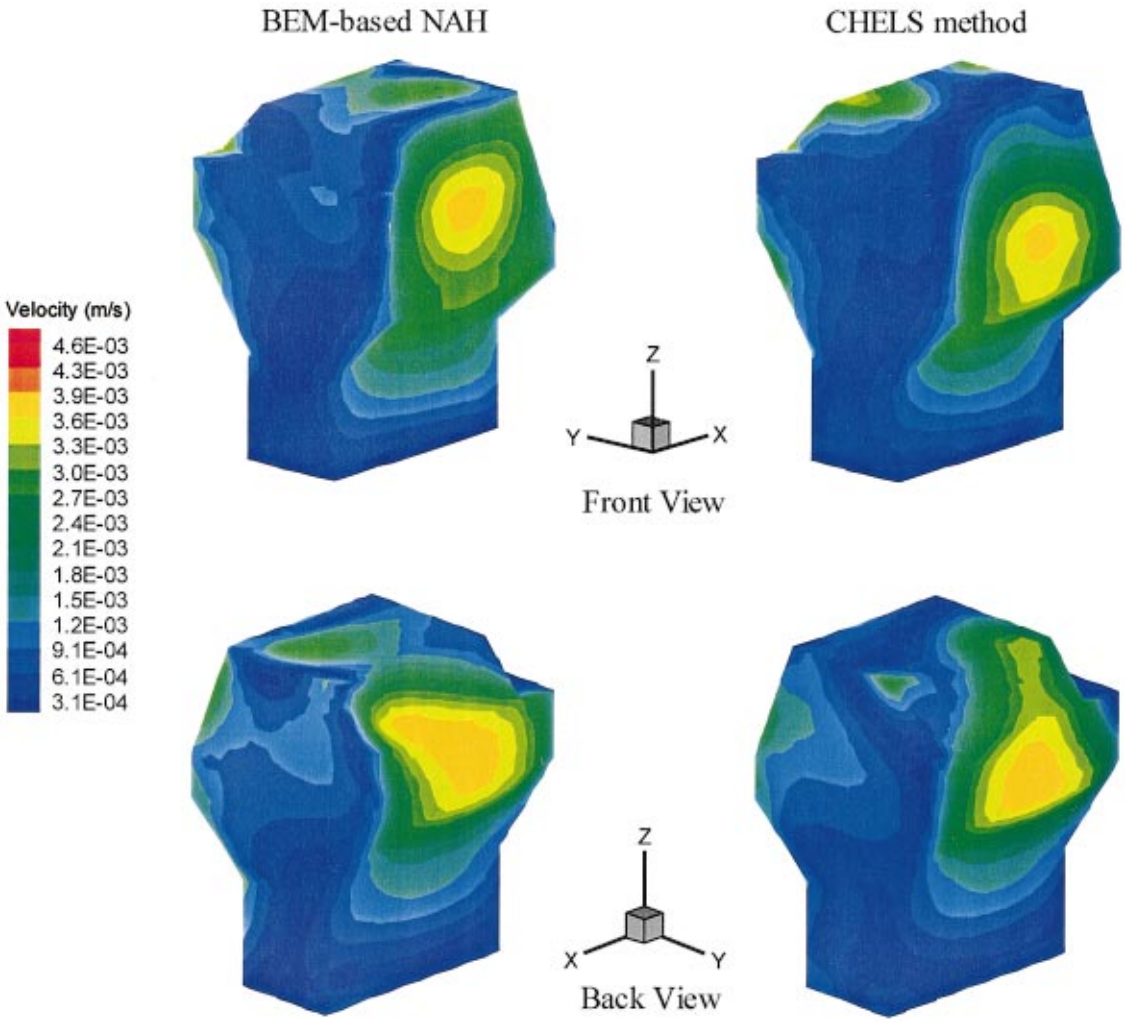


FIG. 8. Comparison of the reconstructed normal component of surface velocity distributions over the entire surface of the engine block using the CHELS method (right column) and BEM-based NAH (left column) at $ka=1$.

nodes. However, if the measurements are too few, then aliasing may occur and the resulting reconstruction may be distorted. Figure 4 illustrates that when 152 field acoustic pressures are taken as input in the BEM-based NAH for a surface with 602 discrete nodes, the resulting reconstruction of the surface acoustic pressure is severely distorted.

To show the effect of measurement distances on the reconstruction accuracy, we present the reconstructed surface acoustic quantities based on conformal measurements taken at different radial distances $r=0.105$, 0.110, 0.125, and 0.150 m under $ka=1.46$. Figure 5 depicts that as measurement distances increase, more evanescent waves are lost. As a result, the reconstructed normal surface velocity is distorted. However, the accuracy of reconstruction of surface acoustic pressure remains essentially unchanged (results omitted for brevity). This is because the normal surface velocity contains more near-field effects than the surface acoustic pressure does. The same phenomenon is observed at higher frequencies (results omitted for brevity). These results clearly demonstrate the importance of measurements in the near field.

B. Nonseparable geometry

The second example is a simplified engine block with an overall length of 0.460 m, overall width of 0.435 m, and overall height of 0.630 m. To test the effectiveness of the CHELS method, sharp edges and corners and abrupt changes in surface contours are built in this model. Analytic solutions for this object do not exist and numerical solutions must be sought.

To simulate acoustic radiation from this engine block in a free field, harmonic excitations of different amplitudes are assumed on three arbitrarily selected surfaces: 5×10^5 N/m^2 on the top and 2×10^5 N/m^2 on part of the front and back surfaces at various frequencies (see Fig. 6). The bottom of the block is clamped with zero displacement and slope, and the rest surfaces are unconstrained. The normal surface velocity distributions are obtained using the standard FEM codes and the surface acoustic pressures are determined using the BEM codes with 1548 triangular elements and 776 nodes. Once the surface acoustic quantities are specified, the field acoustic pressures can be calculated using the Helmholtz integral formulation.

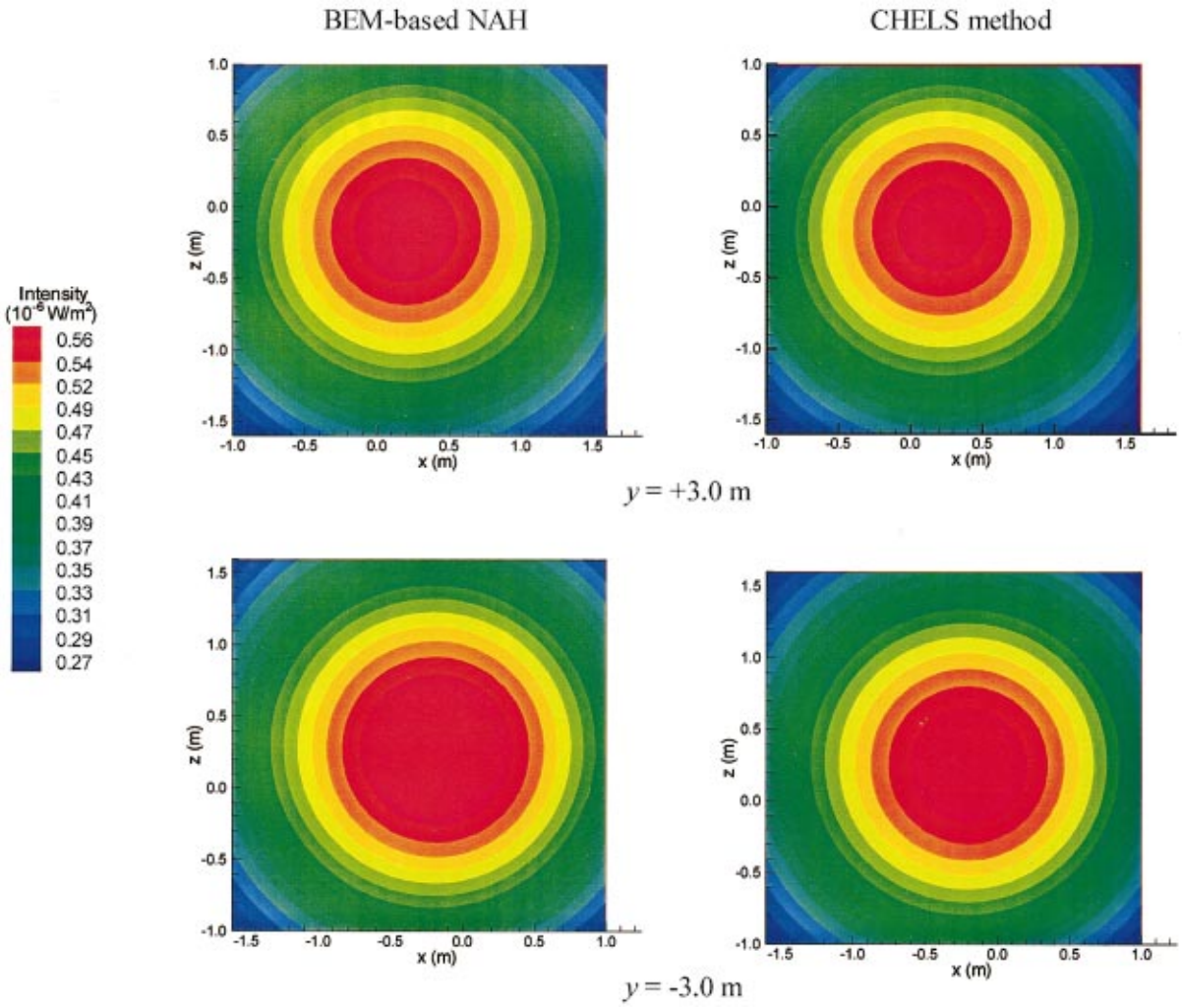


FIG. 9. Comparison of the reconstructed normal component of the time-averaged acoustic intensity distribution over two planes at $y=\pm 3.0$ m measured with respect to the center of the engine block using the CHELS method (right column) and BEM-based NAH (left column) at $ka=1$.

To reconstruct the surface acoustic quantities via the CHELS method, we take $M=277$ measurements over an imaginary sphere of radius $r=0.408$ m that encloses the engine block. In particular, we use $M_1=56$ as input to determine the coefficients and the rest for optimization. The optimal expansion number in this case is roughly $J_{\text{op}}=22$ for the frequencies considered. Once this is done, the field acoustic pressures are regenerated by Eq. (11) and used as input to reconstruct the surface acoustic quantities.

Note that one can always divide the measurements into two groups of any sizes, one for input and the other for optimization. Since in the CHELS method measurements are taken over a spherical surface rather than a conformal surface, the clearance between the measurement and source surfaces may vary from one point to another. As a result, the amounts of the evanescent waves captured in the measurements may be different. Thus, one need to find a low-pass filter to eliminate the evanescent waves that drop below the background noise level, which is equivalent to specifying the optimal expansion number J_{op}. Since in this case the accuracy of the input data changes with the measurement positions, it is better to select a smaller value of M_1 and a larger value of M_2.

It is emphasized that theoretically the expansion solutions (1) using the spherical Hankel functions and spherical harmonics are valid for a spherical surface. For a nonspherical surface, these expansion solutions are approximate.[19] Hence it is not realistic to expect the reconstructed surface acoustic quantities for a non-spherical object to be as good as that for a spherical surface.

Also noted is that the accuracy of reconstruction varies with frequency. To ensure a near-field measurement, the standoff distance d must satisfy the three conditions simultaneously:[2] (1) $d \ll a$; (2) $d \ll \lambda_{\min}=c/f_{\max}$; and (3) $d \ll k_{\max}a^2/2$, where $k_{\max}=2\pi/\lambda_{\min}$ and a is the source characteristic dimension. The first condition can be satisfied by taking measurements at very close range. The second and third conditions require the frequency to be neither very high nor very low. Thus for a source of dimension a and fixed standoff distance d the dimensionless frequency should fall between $2d/a \ll ka \ll 2\pi a/d$. This requirement holds true for all NAH, including the CHELS method. Since in the CHELS

method the standoff distance d is not constant, we should reduce the frequency range by using $2d_{\max}/a \ll ka \ll 2\pi a/d_{\max}$, where $d_{\max}$ is the maximum clearance between the measurement and source surfaces.

In this case the average dimension of the engine is $(0.435+0.460+0.630)/3=0.508$ m and the maximum clearance between measurement and source surface is $d_{\max}=0.191$ m. So we should limit reconstruction to the dimensionless frequency range of $0.752 \ll ka \ll 16.7$.

In what follows, we present the reconstructed acoustic field on the engine block surface. For validation purpose, we use Eq. (10) to reconstruct the surface acoustic quantities under the same $N=776$ input data. Figures 7 and 8 depict the comparisons of the reconstructed acoustic pressures and normal velocities on the surface of the engine block via the CHELS method and the BEM-based NAH, respectively, at $ka=1$.

It is emphasized that this engine block represents a very complex structure, containing sharp edges and corners and abrupt changes in surface contours. Yet satisfactory agreements between the CHELS method and BEM-based NAH are obtained in all cases.

Equation (12) can be used to reconstruct acoustic radiation in the far field as well. This is straightforward because all the points are now outside the minimum sphere so the acoustic field are adequately represented by the outgoing spherical waves. Moreover, the loss of the evanescent waves has a negligible impact on the resultant reconstruction of the radiated acoustic field, just like near-field acoustic scaning.[25]

Figure 9 shows the comparison of the reconstructed normal component of the time-averaged acoustic intensity using the CHELS method on two planar surfaces (2.6×2.6 m^2) at $y=\pm3$ m measured from the center of the engine block vs the BEM results. Note that the peak amplitude of the time-averaged intensity on the front plane is slightly lower than that of the back plane. This is because the area on which the harmonic force exerted on the front is slightly lower than that on the back of the engine block. This result indicates that one can use NAH to locate the excitation acting on an object that is responsible for acoustic radiation.

V. CONCLUSIONS

A combined HELS method and BEM-based NAH, known as the CHELS method is developed for reconstructing acoustic radiation from an arbitrarily shaped object. For the same amount of measurements taken on or beyond a spherical surface that encloses the object, the CHELS method may give a more accurate reconstruction by increasing the number of discrete nodes and the necessary input data than the BEM-based NAH does. Since these input data are calculated but not measured, the reconstruction efficiency is greatly enhanced. However, since the measurements are taken over a spherical surface rather than a conformal surface, the accuracy of reconstruction of the surface acoustic quantities may be limited because some evanescent waves may be lost. On the other hand, if reconstruction of acoustic radiation in the far field is desired where the evanescent waves are negligibly small, the CHELS method can be straightforward and much more efficient than BEM.

ACKNOWLEDGMENT

This work was supported by a grant from National Science Foundation, Grant No. ITR/ACS 0081270.

[1] 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).
[2] S. F. Wu, "On reconstruction of acoustic pressure fields using the Helmholtz equation least squares method," J. Acoust. Soc. Am. **107**, 2511–2522 (2000).
[3] N. 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).
[4] 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).
[5] S. F. Wu and J. Yu, "Reconstructing interior acoustic pressure fields via Helmholtz equation-least squares method," J. Acoust. Soc. Am. **104**, 2054–2060 (1998).
[6] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981), Chap. 4, pp. 165–194.
[7] W. A. Veronesi and J. D. Maynard, "Digital holographic reconstruction of sources with arbitrarily shaped surfaces," J. Acoust. Soc. Am. **85**, 588–598 (1989).
[8] G.-T. Kim and B.-T. Lee, "3-D sound source reconstruction and field reprediction using the Helmholtz integral equation," J. Sound Vib. **136**, 245–261 (1990).
[9] 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).
[10] S.-C. Kang and J.-G. Ih, "The use of partially measured source data in near-field acoustical holography based on the BEM," J. Acoust. Soc. Am. **107**, 2472–2479 (2000).
[11] B.-K. Kim and J.-G. Ih, "On the reconstruction of the vibro-acoustic field over the surface enclosing an interior space using the boundary element method," J. Acoust. Soc. Am. **100**, 3003–3016 (1996).
[12] B. K. Gardner and R. J. Bernhard, "A noise source identification technique using an inverse Helmholtz integral equation method," Trans. ASME, J. Vib., Acoust., Stress, Reliab. Des. **110**, 84–90 (1988).
[13] E. G. Williams, B. H. Houson, P. C. Herdic, S. T. Raveendra, and B. Gardner, "Interior near-field acoustical holography in flight," J. Acoust. Soc. Am. **108**, 1451–1463 (2000).
[14] H. A. Schenck, "Improved integral formulation for acoustic radiation problems," J. Acoust. Soc. Am. **44**, 41–58 (1968).
[15] S.-C. Kang and J.-G. Ih, "Use of nonsingular boundary integral formulation for reducing errors due to near-field measurements in the boundary element method based near-field acoustic holography," J. Acoust. Soc. Am. **109**, 1320–1328 (2001).
[16] B.-K. Kim and J.-G. Ih, "Design of an optimal wave-vector filter for enhancing the resolution of reconstructed source field by near-field acoustical holography (NAH)," J. Acoust. Soc. Am. **107**, 3289–3297 (2000).
[17] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, San Diego, California, 1999).
[18] P. J. Davis and P. Rabinowitz, "Advances in orthonormalizing computation," in *Advances in Computers*, edited by F. L. Alt (Academic, New York, London, 1961), Vol. 2, pp. 55–133.
[19] V. Isokov and S. F. Wu (unpublished).
[20] N. E. Rayess and S. F. Wu, "Why can spherical harmonics be used to describe the sound field generated by nonspherical sources?" J. Acoust. Soc. Am. **108**, 2504 (2000).
[21] E. G. Williams, "Regularization methods for near-field acoustical holography," J. Acoust. Soc. Am. **110**, 1976–1988 (2001).

[22] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems* (SIAM, Philadelphia, PA, 1998).
[23] B.-D. Lim and S. F. Wu, ''Determination of the optimal number of expansion terms in the HELS method,'' J. Acoust. Soc. Am. **108**, 2505 (2000).
[24] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1986).
[25] T. B. Hansen, ''Spherical expansions of time-domain acoustic fields: Application to near-field scanning,'' J. Acoust. Soc. Am. **98**, 1204–1215 (1995).

Null-broadening in a waveguide

J. S. Kim,[a] W. S. Hodgkiss, W. A. Kuperman, and H. C. Song
Marine Physical Laboratory/Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0238

(Received 1 March 2001; revised 25 March 2002; accepted 20 April 2002)

Null-broadening, introduced in plane wave beamforming, is extended to an ocean waveguide in the context of matched field processing. The method is based on the minimum variance processor with white noise constraint and the distribution of fictitious sources using the theory of waveguide invariants. The proposed method is demonstrated in simulation as well as with data collected during the SWellEx-96 experiment. As another application, it is shown that the width of a null can be controlled in an adaptive time reversal mirror with a source-receive array. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1488139]

PACS numbers: 43.60.Gk, 43.30.Wi, 43.20.Fn [SAC]

I. INTRODUCTION

Controlling the shape of a spatial filter is a useful technique in transmission and reception with an adaptive array. The adaptive beam pattern is generated via adaptive weights on the array, for example, to broaden the main lobe or a null as well as to reduce the sidelobes of the spatial filter. Recently, null-broadening in plane wave beamforming (PBF) has been introduced by Mailloux[1] and Zatman[2] and Guerci[3] derived a covariance matrix taper (CMT) function to broaden a null. In the current paper, null-broadening is extended to a waveguide utilizing the theory of waveguide invariants.

Null-broadening provides robust adaptive nulling in passive array signal processing when the sources of interference are confined to either a specific direction (plane wave beamforming) or a specific cell (matched field processing). Moreover, in the transmission of acoustic energy, the null-broadening technique provides robust null-steering and control of the null-width in a temporally fluctuating ocean environment. Therefore, the objective of this paper is to develop a method to broaden nulls in an ocean waveguide in the context of matched field processing (MFP).

In order to broaden a null in a waveguide, the theory of waveguide invariants[4–7] is utilized to augment the fictitious interferers in the vicinity of a true interferer. When the minimum variance processor with white noise constraint is applied to minimize the power coming from other than the look-direction, nulls are placed in the fictitious interferer locations so as to broaden the null. The method is applied here to the ocean waveguide to demonstrate null-broadening in simulation and with ocean data. As another application, the method is employed to broaden the null steered by an adaptive time reversal mirror.[8]

In Sec. II, the null-broadening method in plane wave beamforming is reviewed and the implication of using short-length fast Fourier transforms (FFT's) to find adaptive weights is discussed. The null-broadening method in a waveguide is developed and realized via simulation in Sec. III. In Sec. IV, ocean acoustic data is analyzed to show the effect of null-broadening in matched field processing. As another application, null-broadening with an adaptive time reversal mirror (ATRM) is demonstrated in Sec. V.

II. NULL-BROADENING IN PLANE WAVE BEAMFORMING: FREE SPACE

In plane wave beamforming, a technique to make a null broader by tapering the cross-spectral density matrix (CSDM) has been developed by augmenting the CSDM with multiple interferers in the vicinity of the true interferer location[1] and by dispersion synthesis.[2] These two approaches use different formulations but are based on the same principle of distributing the additional interferers around the true location of the interferer.

The plane wave beamforming derivations which are simple and instrumental in extending their approach to matched field processing (MFP) in an ocean waveguide are review in Sec. II A. The simulation results are presented in Sec. II B.

A. Theory

In plane wave beamforming, the beam pattern of the Bartlett processor is a correlation between a data vector $\mathbf{d}(\theta_s)$ and the search vector (or steering or replica vector) $\mathbf{d}(\theta)$ which is expressed as

$$B_{BT}(\theta)=\mathbf{d}(\theta_s)^{\dagger}\mathbf{d}(\theta), \tag{1}$$

where θ_s is source direction and $(\ )^{\dagger}$ denotes the Hermitian or conjugate transpose. For the minimum variance distortionless response (MVDR) processor, the weight vector $\mathbf{w}(\theta)$ in the direction θ is

$$\mathbf{w}(\theta)=\frac{\mathbf{K}^{-1}\mathbf{d}(\theta)}{\mathbf{d}^{\dagger}(\theta)\mathbf{K}^{-1}\mathbf{d}(\theta)}, \tag{2}$$

where $\mathbf{K}$ is the cross-spectral density matrix (CSDM) defined as $\mathbf{K}=\Sigma_{\text{source}}\mathbf{d}(\theta_s)\mathbf{d}^{\dagger}(\theta_s)$ assuming the sources are uncorrelated for the moment and ignoring any diffuse or uncorrelated noise at the array. The beam pattern of the MVDR processor while looking at θ_l is

$$B_{\text{MVDR}}(\theta)=\mathbf{w}^{\dagger}(\theta_l)\mathbf{d}(\theta). \tag{3}$$

[a]The work was done during the author's sabbatical leave at MPL. Permanent affiliation: Korea Maritime University, Pusan 606-791, Korea.

The distortionless response in the look-direction requires that

$$B_{\mathrm{MVDR}}(\theta_l)=\mathbf{w}^{\dagger}(\theta_l)\mathbf{d}(\theta_l)=1. \tag{4}$$

The CSDM for plane waves coming from $u_l=\sin\theta_l$, where θ_l is measured from the array normal, can be written as

$$k_{mn}=\sigma_l^2 e^{-j(2\pi/\lambda)(x_m-x_n)u_l}, \tag{5}$$

where σ^2 is the source strength, x_m and x_n represent the location of the mth and nth array elements, $\lambda=c/f$ denotes the wavelength, c is the speed of propagation, and f is frequency.

In order to obtain multiple nulls centered around the direction u_l, Mailloux[1] distributed plane waves coming from N directions as

$$u=u_l+q\,\delta, \quad \text{for } -\frac{N-1}{2}\leqslant q\leqslant\frac{N-1}{2}, \tag{6}$$

where δ is defined as $\delta=W/(N-1)$ to obtain a trough width of W between the outermost nulls. The summation over q can be written as

$$K_{mn}=\sum_q k_{mn}=\sum_q \sigma_l^2 e^{j(2\pi/\lambda)(x_m-x_n)(u_l+q\delta)}$$

$$=\frac{\sin(N\Delta)}{\sin(\Delta)}k_{mn}, \tag{7}$$

where $\Delta=\pi(x_m-x_n)\delta/\lambda$.

Zatman[2] distributed the fictitious sources in a different way. By slightly changing the frequency and keeping the incident direction the same, the array sees the data vector as one coming from a different direction. Thus, the augmentation of the fictitious sources can be achieved by integrating the CSDM over a desired frequency bandwidth

$$K_{mn}=\frac{1}{b_w}\int_{f-b_w/2}^{f+b_w/2}k_{mn}\,df=\frac{\sin(\pi b_w\tau_{mn})}{\pi b_w\tau_{mn}}k_{mn}, \tag{8}$$

where $\tau_{mn}=(x_m-x_n)u_l/c$ is the time delay between elements. Note that Mailloux's formulation directly places the fictitious sources in angular space, while Zatman's formulation uses a frequency perturbed CSDM.

The null-broadening approach can be implemented through the concept of a covariance matrix taper (CMT) where the CSDM is modified by a Hadamard or Schur product tapering function.[3] This is applicable to null-broadening in plane wave beamforming. However, in matched field processing the tapering function is coupled with the individual terms of the mode function so that the tapering function cannot be separated from the CSDM. Fortunately, there exists an invariant property associated with the waveguide that allows us to displace nulls by slightly changing the frequency, thus augmenting the nulls around the true location of the interferer that is to be nulled (the details are discussed in Sec. III A).

B. Plane wave beamforming simulation

The elements of the ideal CSDM with uncorrelated noise can be written[9] as

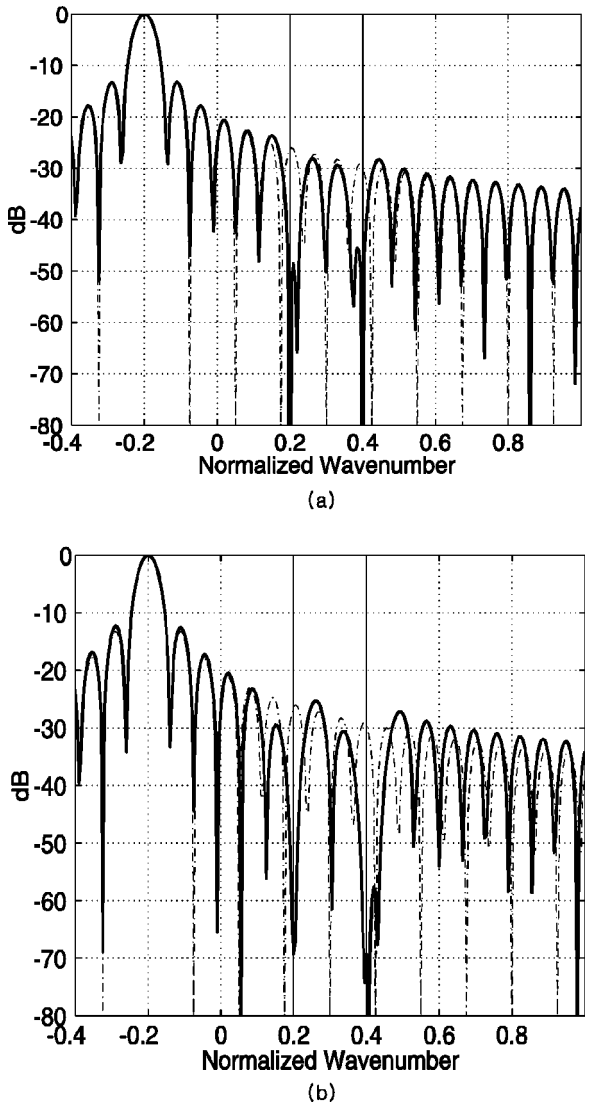


FIG. 1. Beam pattern of the WNC processor for PBF (a) without null-broadening and (b) with null-broadening. The broken line is the beam pattern for a Bartlett processor.

$$k_{mn}=\sum_{l=1}^{3}\sigma_l^2 e^{-j(2\pi\Delta u_l/\lambda)(m-n)}+\sigma_w^2\delta_{mn}. \tag{9}$$

For simulation, one source with $\sigma_1^2=6$ dB, $u_1=-0.2$ and two interferers with $\sigma_2^2=36$ dB, $u_2=0.2$ and $\sigma_3^2=36$ dB, $u_3=0.4$ are used, where u is defined as $u=\sin\theta$. The uncorrelated noise level is $\sigma_w^2=0$ dB. The total number of array elements is $N=64$ and the elements are separated by $\lambda/4$, where λ is the wavelength.

In Fig. 1(a), the beam pattern is shown for the exact CSDM in Eq. (9) while looking at the target at the normalized wave number $u_1=\sin\theta_1=-0.2$. The broken line is the beam pattern of the Bartlett processor and the solid line is that of the MVDR processor with white noise gain constraint (WNC).[10–12] Throughout the paper, -6 dB constraint is used for the WNC processor. The two vertical lines denote the direction of interferers at $u_2=\sin\theta_2=0.2$, and at $u_2=\sin\theta_2=0.4$. It is noted that the WNC beam pattern exhibits nulls in the direction of the interferers. The beam pattern shown in Fig. 1(b) was obtained using the integrated CSDM defined in

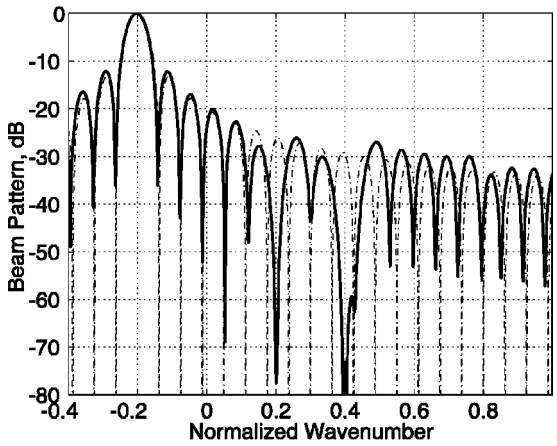


FIG. 2. Beam pattern with 256 point FFTs and 256 snapshots. The null-broadening is achieved by using short-length FFTs. The broken line is the beam pattern for a Bartlett processor.

Eq. (8) where $f=100$ Hz and $b_w=5$ Hz. In this case, these results were computed using a 4096-point FFT on independently generated frequency components with frequency resolution of $\Delta f=1/(4096\Delta t)$. Since the sampling rate was 1500 samples/second, $\Delta f=0.3662$ Hz. When the frequency is perturbed, the position of the null seen from the array is calculated from the relation

$$\frac{2\pi}{\lambda}u=\text{const}, \tag{10}$$

which leads to

$$u'=\frac{f}{f'}u \quad \text{or} \quad \frac{\sin\theta'}{\lambda'}=\frac{\sin\theta}{\lambda}, \tag{11}$$

where λ' is the wavelength of the perturbed frequency and, u' is the corresponding null location. It is shown in Fig. 1(a) that the null of the beam pattern is broadened through the use of wideband CSDM. In addition, it is noted that null-broadening is more significant for the source at a larger angle since the width of angular augmentation becomes larger for a larger arrival angle, for a given frequency bandwidth.

C. Effect of using short-length FFTs

When the number of snapshots is limited due to observing a fast moving source, a short-length FFT can be used to increase the number of snapshots involved in estimating the CSDM. In this section, it is shown that the use of a short-length FFT produces the same effect as using the CSDM integrated over frequency and results in null-broadening.

Figure 2 shows the beam pattern obtained using 256 snapshots with 256 point FFTs. The frequency resolution is now $\Delta f=1/(256\Delta t)=5.869$ Hz so that a single FFT bin is as wide as the averaging bandwidth represented in Fig. 1(b). In this case, the CSDM components for the source and two interferers were estimated by averaging outer products of broadband array signal vectors (snapshots). These three components plus uncorrelated noise then were added together to form the CSDM estimate.

III. NULL-BROADENING IN MATCHED FIELD PROCESSING: OCEAN WAVEGUIDE

In this section, a null-broadening algorithm in an ocean waveguide is described in the context of matched field processing.

A beam pattern in plane wave beamforming (PBF) can be viewed as a correlation between a signal vector coming from the look-direction and a family of array steering vectors. When this concept is applied to matched field processing (MFP), the equivalent beam pattern is defined as a correlation between a signal vector coming from the look location and a family of array replica vectors. Thus the MFP replica vectors are a generalized form of the steering vectors in PBF.

In Sec. III A, the null-broadening algorithm is described and then simulation results are shown in Sec. III B.

A. Theory

Unlike null-broadening in plane wave beamforming, the covariance matrix taper (CMT) is not explicitly available for null-broadening of matched field processing in a waveguide. From the perspective of mode theory, this can be attributed to the fact that source depth and range are coupled to the mode sum involving a wavenumber spectrum. However, the theory of waveguide invariants[4–7] can be utilized to derive the signal vectors from adjacent ranges at the same depth when the signal vector at the probe source or the interferer location is known. Thus, the distribution of the interferer around the true interferer location can be achieved. The adaptive minimum variance processor with a white noise constraint will then place nulls around the location of the true interferer and broaden the null.

From the theory of waveguide invariants, signal vectors in the vicinity of the interferer and at the same depth are simply a frequency-shifted signal vector coming from the interferer location. Here, we briefly state the practical form of the theory of waveguide invariants applicable to null-broadening in a waveguide.

In a dispersive and multi-modal waveguide, the lines of constant intensity lead to a constant slope between the certain parameters of the waveguide. The invariant, denoted as β, characterizes the relation between the range r and angular frequency ω as

$$\beta=\frac{r}{\omega}\frac{\delta\omega}{\delta r}. \tag{12}$$

The parameter β defined in Eq. (12) is known to be approximately invariant with respect to range and is equal to one in a Pekeris waveguide. For a Pekeris waveguide, Eq. (12) can be rewritten to calculate the new range r' in terms of new frequency ω':

$$r'=\frac{\omega'}{\omega}r. \tag{13}$$

Equation (13) states that the acoustic field at (r,ω') approximates the value at (r',ω). Therefore, broadband data received at an array can be used to estimate the CSDM in the vicinity of the source range at the same depth so as to place

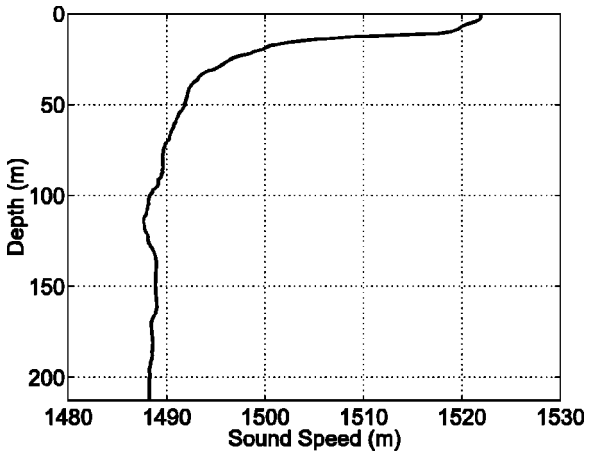


FIG. 3. Sound speed profile for simulation and SWellEx-96 experiment.

a null without measuring the Green's function from fictitious null positions.

B. Simulation with SWellEx-96 environment

The sound profile as a function of depth for our simulation is shown in Fig. 3 which is the same environment as SWellEx-96 experiment.[13] The trajectories of target and interferer assume the same situation as the experiment for the purpose of comparison (see Sec. IV B). The target is at a depth of 20 m and at a range of 5040 m, while the interferer to be nulled is at the same depth and at a range of 3310 m. The radiated power from the interferer is assumed to be 6 dB higher than the radiation from the target.

Since the applicability of the null-broadening technique is dependent on the validity of waveguide invariant theory and the width of the null is determined from the value of β, the transmission loss versus frequency is plotted in Fig. 4. This figure shows that β in Eq. (12) is approximately one as in a Pekeris waveguide.

The simulation results with the ideal CSDM in Fig. 5 display the null-broadening effect at 62 Hz that is to be compared later with the beam pattern from SWellEx-96 data in Fig. 10. Figure 5(a) is for a Pekeris waveguide with the same depth as in SWellEx-96. The broken line is the beam pattern for the conventional matched field processor, the solid line is the beam pattern for the minimum variance processor with white noise constraint (WNC), and the thick solid line is for the same processor as the solid line with null-broadening. In the simulation, since the same sampling parameters are used as in the data analysis, the frequency resolution is given as $\Delta f = 1/(\Delta t * 4096) = 0.3662$ Hz, where $\Delta t = 1/1500$. From Eq. (13), the range shift in a range-independent environment caused by a 7-bin frequency shift is related by the following equation

$$r' = \frac{\omega'}{\omega} r = \frac{2\pi(62 \pm 0.3662 \times 7)}{2\pi \times 62} 3310 \simeq 3310 \pm 140 \text{ m}.$$

The predicted null locations corresponding to the augmented interferers agree well with the null locations shown in Fig. 5(a).

Figure 5(b) shows the nulling pattern as a function of range at a depth of 20 m for the SwellEx-96 experiment. As

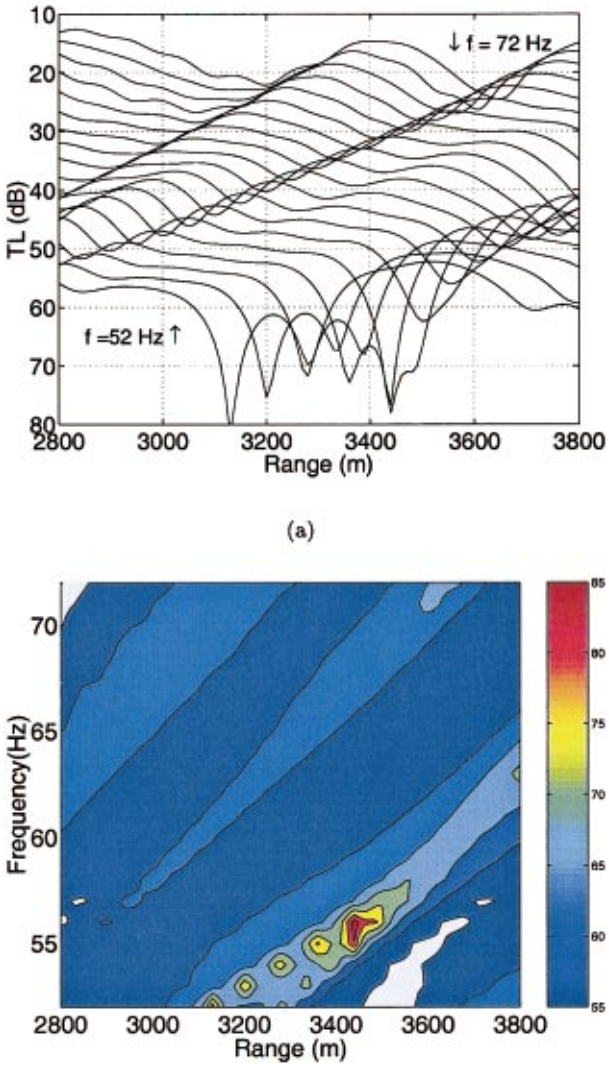


FIG. 4. Waveguide invariant in SWellEx-96. (a) Transmission loss plot (dB) for 52–72 Hz with a 1-Hz interval. (b) Contour plot of (a).

in Fig. 5(a), the broken line is the beam pattern of the Bartlett processor, the solid line represents the beam pattern of WNC without null-broadening, and the thick solid line is WNC with null-broadening. The figure shows that the Bartlett processor is not aware of the interferer, and the null is independent of the location of interferer. However, the beam pattern of WNC places a null at the interferer location $r = 3310$ m.

IV. DEMONSTRATION OF NULL-BROADENING USING OCEAN DATA

In this section, the null-broadening technique is applied to SWellEx-96 data to verify the applicability of the technique to matched field processing in a shallow water waveguide.

In order to apply the null-broadening technique developed in Sec. III, the proper data set requires a wideband signal radiated from a target and an interferer. During the SWellEx-96 experiment, broadband noise radiated from a cargo ship passing by the vertical line array was observed.

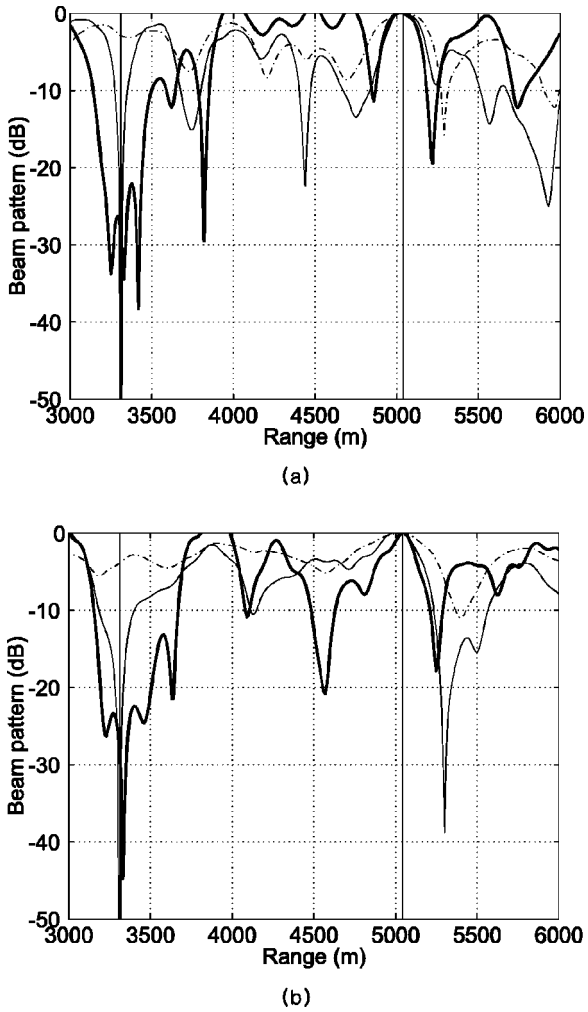


FIG. 5. Simulation with ideal CSDM in (a) Pekeris waveguide and (b) SWellEx-96 environment. The beam pattern of the Bartlett processor is denoted by a broken line, the thin solid line is for WNC without null-broadening, and the thick solid line is for WNC with null-broadening.

Since this data set contains only one broadband noise source as a target, noise from an interferer is simulated by using the data observed at a different time.

In Sec. IV A, the ocean environment and a plan view of the target and interferer are illustrated. The results of null-broadening then are presented in Sec. IV B.

A. Description of SWellEx-96 experiment

The detailed experiment description can be found in Ref. 13. Figure 6(a) shows a plan view of the SwellEx-96 experiment with towed source ship and Fig. 6(b) shows the track of the towed source and interferer. Since the towed source radiates tonals, we are only interested in the broadband noise radiated from the cargo ship. The square marks on the composite radar image shows the track of the cargo ship every 5 min.

B. Two broadband random radiators

In order to apply the null-broadening technique, the following ten frequencies with relatively good signal to noise ratio were used: 53, 54, 56, 59, 60, 62, 66, 69, 72, and 74 Hz. The actual target observed in the data moves from a range of

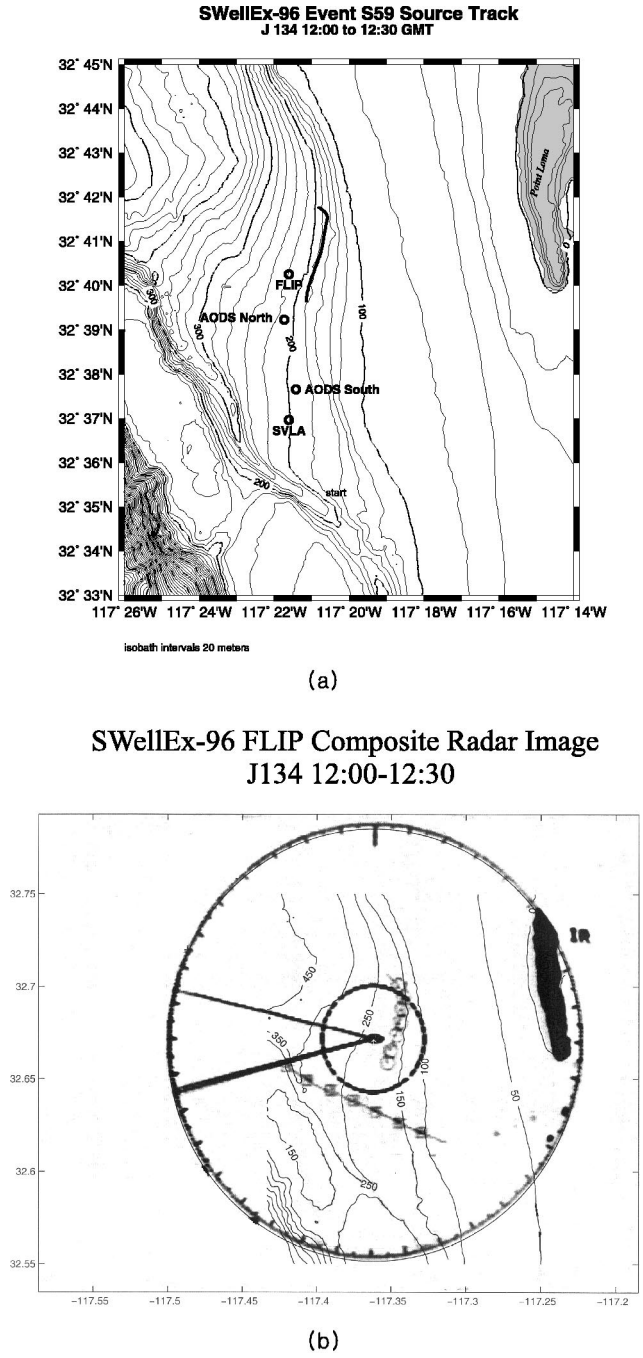


FIG. 6. (a) Plan view of the SWellEx-96 experiment and (b) composite radar image during J134 12:00–12:30.

3200 m at the 14th minute to over 6000 m at the 24th minute. Since we need a two-target situation, the interferer is simulated by superimposing the CSDM of the target between the 14th and 18th minutes elongated twice in time by repeating it twice and reversed in time with 6 dB higher intensity than the target radiation.

In order to track the random radiator, the ambiguity surface at 20-m depth is displayed based on the procedure described in Fig. 7.[14] First, the ambiguity surfaces for range versus depth are constructed for each frequency and the surfaces are incoherently summed over the ten frequency components. Second, this frequency-averaged ambiguity surface

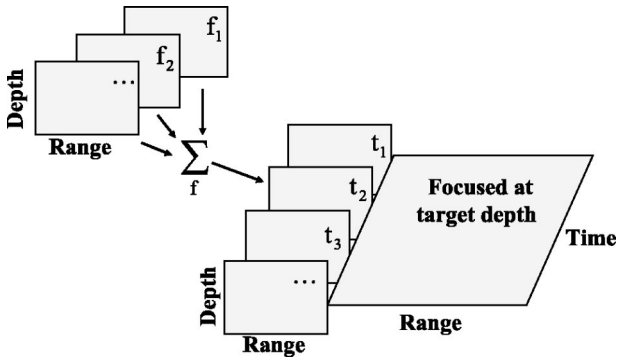


FIG. 7. Construction of ambiguity surface at fixed depth as a function of time and range.

is generated for every time step and is displayed at the selected depth where the target or interferer is located.

Figure 8(a) shows the Bartlett processor ambiguity surface at 20-m depth and Fig. 8(b) is the result from the WNC processor. The interferer trajectory superimposed with a line (running from a range of 3900 m at the 14th minute to 3200 m at the 24th minute) is clearly visible in both figures. However, the ambiguity surface from the Bartlett processor has higher sidelobes so that the trajectory of the target (running from a range of 3200 m at the 14th minute to 6000 m at the 24th minute) is barely visible. The better performance of the WNC processor than the Bartlett processor is attributed to the nulling of sources of interference as shown in Fig. 9.

The beampattern in Fig. 9(a) shows the null along the interferer track while looking at the target. Figure 9(b) shows the null with the null-broadening method using CSDM's averaged across 15 frequency bins centered on each of the frequency components. Note that the null along the trajectory of the interferer (running from a range of 3900 m at the 14th minute to 3200 m at the 24th minute) is wider as a result of the null-broadening method.

A slice of a typical nulling pattern at 62 Hz is shown in Fig. 10, which should be compared to the simulated beam pattern in Fig. 5 calculated at the same frequency. The thin line represents the nulling pattern of WNC without null-broadening and the thick line is with null-broadening where the CSDM has been averaged across 15 frequency bins.

The results at 62 Hz exhibit a somewhat fluctuating nulling pattern. Nonetheless, the frequency-averaged beam pattern in Fig. 9 shows that the null-broadening technique has been applied successfully to SWellEx-96 experimental data.

V. APPLICATION TO NULL STEERING

In this section, null-broadening in an adaptive time reversal mirror (ATRM) is discussed as an example of this technique applied to active acoustics.

The time-reversal mirror (TRM) uses the received signal from a probe source to refocus the signal at the probe source location by back-propagating the time-reversed version of the received signal. The time reversal mirror has been demonstrated in ultrasonics[15,16] and in underwater acoustics.[17] Recently, the adaptive time reversal mirror (ATRM) has been applied to steer the nulls in simulation in an ocean waveguide.[8] Here, we apply the null-broadening technique to the ATRM.

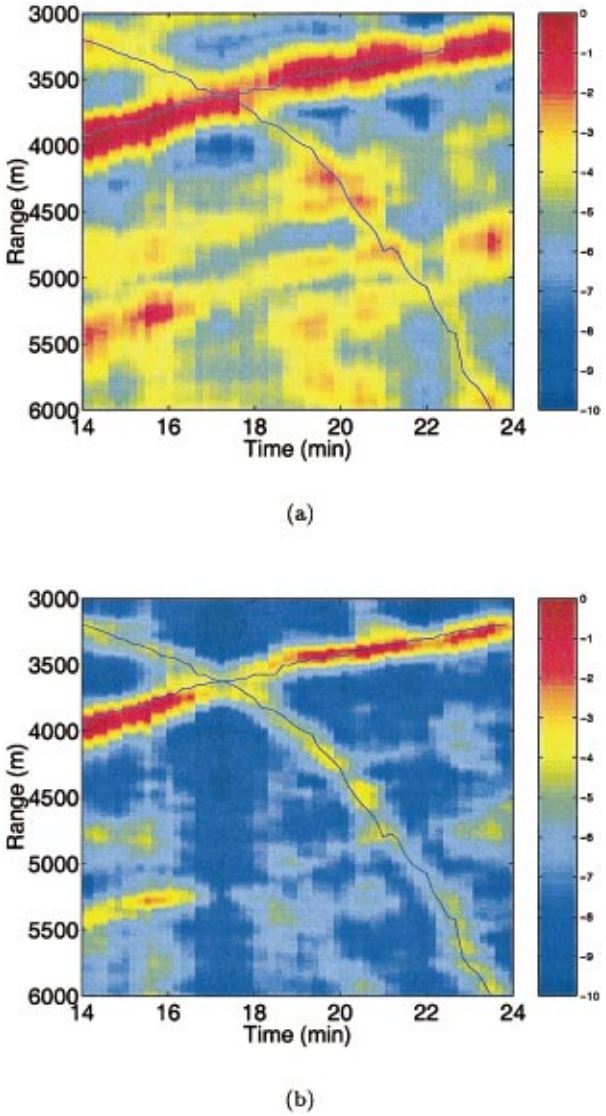


FIG. 8. Trajectories of two broadband sources at 20 m depth. (a) Power (dB) output of the Bartlett processor and (b) the WNC processor without null-broadening.

In Sec. V A, the theory of ATRM is reviewed and the simulation is discussed in Sec. V B.

A. Theory of the adaptive time reversal mirror

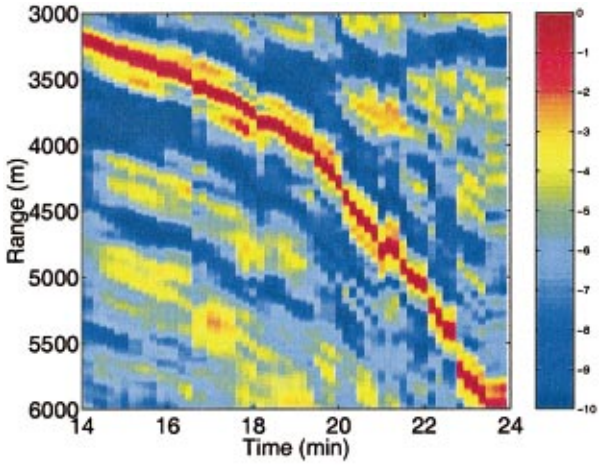
The phase-conjugate field at the field location $\vec{r}$ is written as[8]

$$p(\vec{r}) = \sum_{i=1}^{N} w_i^* g(\vec{r}|\vec{r}_i) = \mathbf{w}^\dagger \mathbf{g}(\vec{r}|\mathbf{r}_{\text{array}}), \tag{14}$$

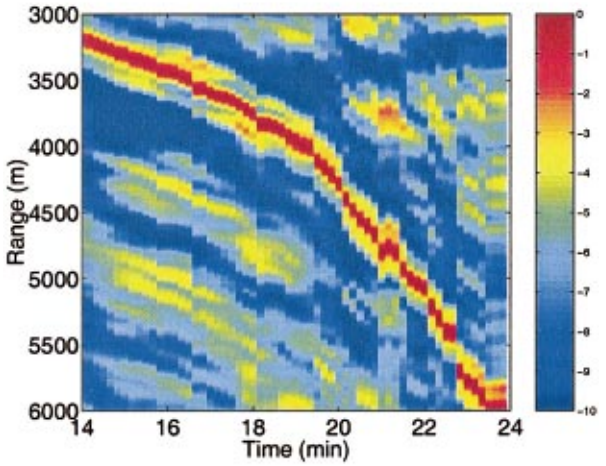
where $\mathbf{r}_{\text{array}}$ is a column vector of array element locations and $\mathbf{g}$ is a column vector of Green's functions. The signal vector $\mathbf{w}$ reduces to $\mathbf{g}$ in a conventional time-reversal mirror.

The ATRM is accomplished by imposing two conditions on the signal vector $\mathbf{w}$ and intensity minimization:

$$\min_{\mathbf{w}} \mathbf{w}^\dagger \mathbf{K} \mathbf{w}, \tag{15}$$



(a)



(b)

FIG. 9. Beam pattern (dB) of (a) the WNC processor without null-broadening and (b) the WNC processor with null-broadening. The beam patterns from ten frequency components are averaged incoherently to produce these plots.

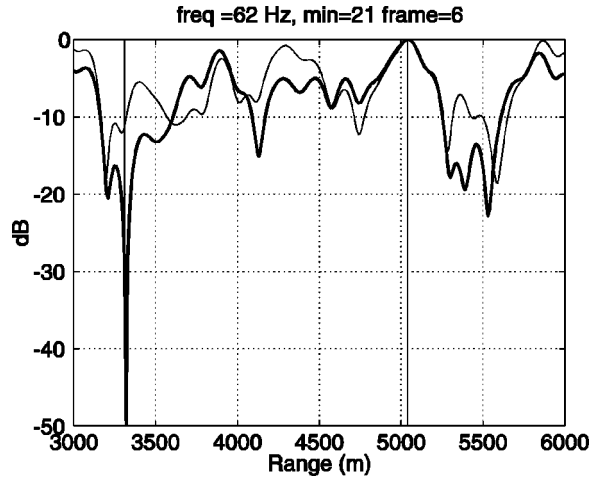
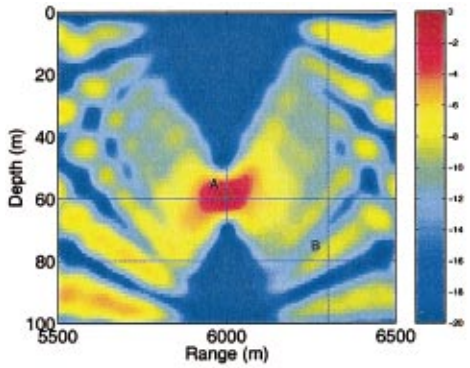
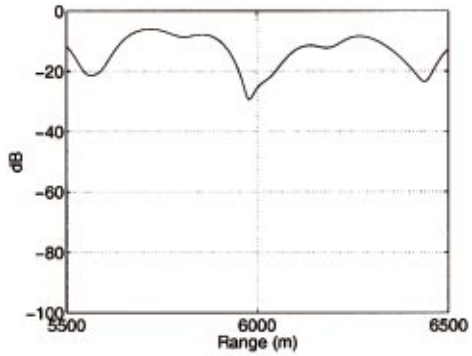


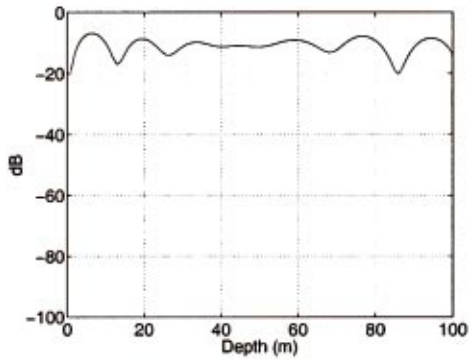
FIG. 10. Beam pattern at 62 Hz just before the 22nd minute at 20-m water depth. The beam pattern of the WNC processor without null-broadening is denoted as a thin solid line and the thick solid line is for the WNC processor with null-broadening.



(a)



(b)



(c)

FIG. 11. (a) Contour of the focused field from the time reversal mirror. Slice of the field passing through the point B: (b) intensity versus range and (c) intensity versus depth. The intensity level is normalized so that the level at the focal point is 0 dB.

and a distortionless response constraint at the focal (probe source) location which can be expressed as

$$\mathbf{w}^{\dagger}\mathbf{g}(\vec{r}_{ps}|\mathbf{r}_{\text{array}})=1. \tag{16}$$

The solution for $\mathbf{w}$ is well known and referred to as the minimum variance method in adaptive array signal processing:[10]

$$\mathbf{w}=\frac{\mathbf{K}^{-1}\mathbf{g}(\vec{r}_{ps}|\mathbf{r}_{\text{array}})}{\mathbf{g}^{\dagger}(\vec{r}_{ps}|\mathbf{r}_{\text{array}})\mathbf{K}^{-1}\mathbf{g}(\vec{r}_{ps}|\mathbf{r}_{\text{array}})}, \tag{17}$$

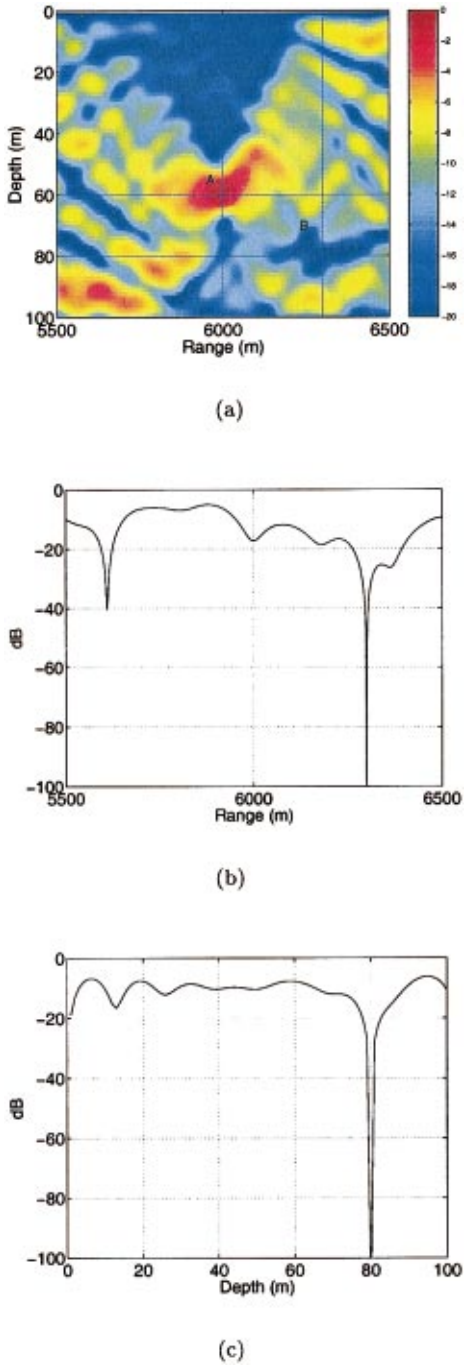


FIG. 12. (a) Contour of the focused field from the time reversal mirror with a steered null at point B. Slice of the field passing through the point B: (b) intensity versus range and (c) intensity versus depth. The intensity level is normalized so that the level at the focal point is 0 dB.

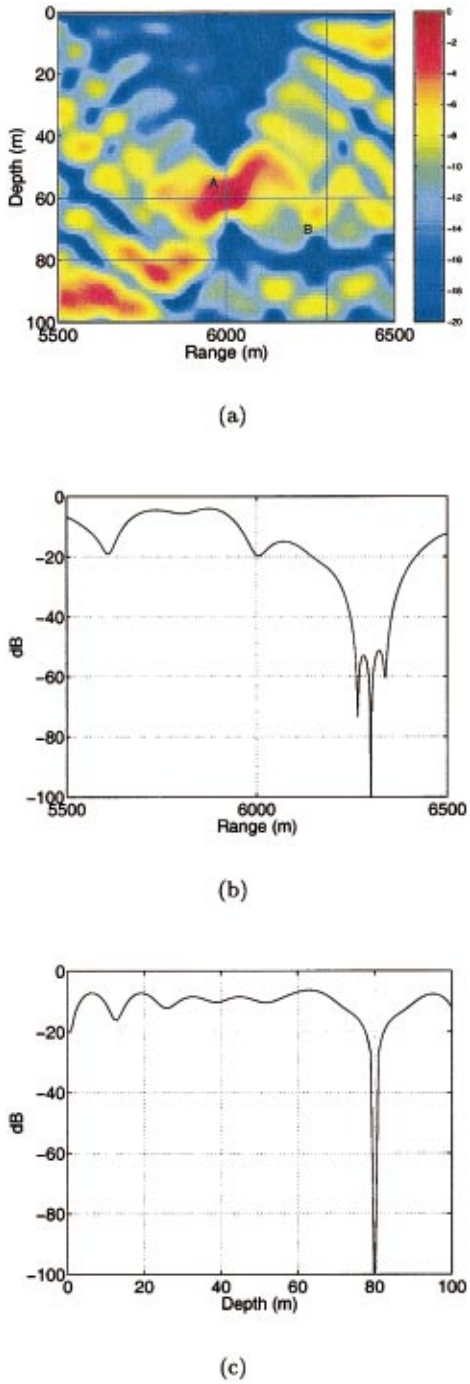


FIG. 13. (a) Contour of the focused field from the adaptive time reversal mirror with a null broadening in range at point B. Slice of the field passing through the point B: (b) intensity versus range and (c) intensity versus depth. The intensity level is normalized so that the level at the focal point is 0 dB.

where **K** is a cross spectral density matrix (CSDM).

When the CSDM is defined as[8]

$$\mathbf{K}=\mathbf{g}(\vec{r}_{ps}|\mathbf{r}_{\text{array}})\mathbf{g}^{\dagger}(\vec{r}_{ps}|\mathbf{r}_{\text{array}})+\mathbf{g}(\vec{r}_{n}|\mathbf{r}_{\text{array}})\mathbf{g}^{\dagger}(\vec{r}_{n}|\mathbf{r}_{\text{array}}), \tag{18}$$

the adaptive signal vector **w** will place a null at $\vec{r}_n$ since the intensity minimization requirement tends to cancel out contributions to the power except at the probe source location. In practice, the signal vector **w** is found from the minimum variance formulation with diagonal loading [e.g., the white noise constraint (WNC) in order to make it robust]. If the placement of nulls is needed at more than one location, additional outer products that correspond to the desired null locations can be added to Eq. (18).

B. Simulation

In this simulation, the same Pekeris waveguide is used as in Sec. III B and in Ref. 8. The Pekeris waveguide is 100 m deep, and the sound speed in the water column is 1500 m/s. The bottom density and sound speed are 1.5 g/cm^3 and 1600 m/s, respectively. We used 300 Hz as the center fre-

quency with a bandwidth of 1.7578 Hz yielding about 37 m null width in range in both directions.

Focusing with the time reversal mirror in a Pekeris waveguide is shown in Fig. 11(a). The level of intensity versus range and depth through point B is shown in Figs. 11(b) and (c), respectively. When a null is steered towards point B with the ATRM, the intensity is as shown in Fig. 12(a) along with horizontal and vertical slices in Figs. 12(b) and (c), respectively. Finally, Fig. 13 illustrates null broadening in range when the null broadening technique is applied to the ATRM.

VI. SUMMARY

In this paper we have described a method to broaden the null of the beam pattern of an adaptive array for reception and transmission in an ocean waveguide based on the theory of waveguide invariants. Null-broadening is demonstrated in matched field processing both in simulation and with data collected during the SWellEx-96 experiment. As another application, the null-broadening technique is demonstrated in an active acoustics example with an adaptive time reversal mirror.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research, Grant No. N00014-94-1-0458. The authors would like to thank James Murray (MPL) for providing the data and related information for the SWellEx-96 experiment.

[1] R. J. Mailloux, ''Covariance matrix augmentation to produce adaptive array pattern troughs,'' Electron. Lett. **31**, 771–772 (1995).
[2] M. Zatman, ''Production of adaptive array troughs by dispersion synthesis,'' Electron. Lett. **31**, 2141–2142 (1995).
[3] J. R. Guerci, ''Theory and Application of Covariance Matrix Tapers for Robust Adaptive Beam-forming,'' IEEE Trans. Signal Process. **47**, 977–985 (1999).
[4] S. D. Chuprov, ''Interference structure of a sound field in a layered ocean,'' in *Acoustics of the Ocean: Current Status* (in Russian), edited by L. M. Brekhovskikh and I. B. Andreevoi (Nauka, Moscow, 1982), pp. 71–91.
[5] G. A. Grachev, ''Theory of acoustic field invariants in layered waveguide,'' Acoust. Phys. **39**, 33–35 (1993).
[6] H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, ''A time-reversal mirror with variable range focusing,'' J. Acoust. Soc. Am. **103**, 3234–3240 (1998).
[7] G. L. D'Spain and W. A. Kuperman, ''Application of waveguide invariants to analysis of spectrograms from shallow water environments that vary in range and azimuth,'' J. Acoust. Soc. Am. **106**, 2454–2468 (1999).
[8] J. S. Kim, H. C. Song, and W. A. Kuperman, ''Adaptive time reversal mirror,'' J. Acoust. Soc. Am. **109**, 1817–1825 (2001).
[9] A. B. Baggeroer and H. Cox, ''Passive Sonar Limits Upon Nulling Multiple Moving Ships with Large Aperture Arrays,'' in *Proceedings of the 33rd Asilomar Conference on Signals, Systems, and Computers*, IEEE Computer Society (1999), pp. 103–108.
[10] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing—Concepts and Techniques*, Prentice Hall Signal Processing Series, Alan V. Oppenheim, Series Editor (Prentice Hall, Englewood Cliffs, NJ, 1993).
[11] H. Cox, ''Robust adaptive beamforming,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**, 1365–1376 (1987).
[12] J. Capon, ''High resolution frequency wavenumber spectrum analysis,'' Proc. IEEE **57**, 1408–1418 (1969).
[13] N. O. Booth, A. T. Abawi, P. W. Schey, and W. S. Hodgkiss, ''Detectability of low-level broad-band signals using adaptive matched-field processing with vertical aperture arrays,'' IEEE J. Ocean. Eng. **25**, 296–313 (2000).
[14] N. Booth, P. Schey, and W. S. Hodgkiss, ''Detection of low-level broadband signals using adaptive matched-field processing,'' J. Acoust. Soc. Am. **101**, 3047 (1997).
[15] M. Fink, ''Time-reversal mirrors,'' J. Phys. D **26**, 1330–1350 (1993).
[16] M. Fink, ''Time-reversed acoustics,'' Phys. Today **50**, 34–40 (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).

Deriving a cochlear transducer function from low-frequency modulation of distortion product otoacoustic emissions

Lin Bian,[a] Mark E. Chertoff, and Emily Miller
Hearing and Speech Department, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, Kansas 66160

(Received 18 December 2001; revised 3 May 2002; accepted 6 May 2002)

In this paper, a new method is introduced to derive a cochlear transducer function from measuring distortion product otoacoustic emissions (DPOAEs). It is shown that the cubic difference tone (CDT, $2f_1-f_2$) is produced from the odd-order terms of a power series that approximates a nonlinear function characterizing cochlear transduction. Exploring the underlying mathematical formulation, it is found that the CDT is proportional to the third derivative of the transduction function when the primary levels are sufficiently small. DPOAEs were measured from nine gerbils in response to two-tone signals biased by a low-frequency tone with different amplitudes. The CDT magnitude was obtained at the peak regions of the bias tone. The results of the experiment demonstrated that the shape of the CDT magnitudes as a function of bias levels was similar to the absolute value of the third derivative of a sigmoidal function. A second-order Boltzmann function was derived from curve fitting the CDT data with an equation that represents the third derivative of the Boltzmann function. Both the CDT-bias function and the derived nonlinear transducer function showed effects of primary levels. The results of the study indicate that the low-frequency modulated DPOAEs can be used to estimate the cochlear transducer function.
[DOI: 10.1121/1.1488943]

PACS numbers: 43.64.Jb, 43.64.Bt, 43.64.Kc [BLM]

I. INTRODUCTION

The inner ear is a nonlinear mechanical transducer that transforms sound pressures into electrical voltages. The cochlear transduction process consists of a forward mechano-electric transduction (MET), and a reverse transduction by which the outer hair cell (OHC) mechanically influences the vibration of the cochlear partition. Cochlear MET can be described by a sigmoid-shaped function, relating the displacement of the basilar membrane (BM) to the electrical response of hair cells [Fig. 1(A)]. Several mathematical functions have been proposed to model the transfer characteristics of MET, e.g., a hyperbolic-tangent function (Weiss and Leong, 1985), a Boltzmann function [$f_B(x)$] of the first-order (Patuzzi and Moleirinho, 1998), and a third-order polynomial function (Chertoff *et al.*, 1996; Bian and Chertoff, 1998). The mechanics of the stereocilial movement and kinetics of the ion channels prompted a three-state model for hair cell transduction (Corey and Hudspeth, 1983; Holton and Hudspeth, 1986). Intracochlear recordings indicated that the hair cell input/output function is asymmetrical to its inflection point (Dallos, 1986; Russell *et al.*, 1986). Therefore, a second-order (three-state) $f_B(x)$ [Eq. (1)], has been adopted to characterize MET (Crawford *et al.*, 1989; Kros *et al.*, 1992):

$$y(x)=\frac{A}{1+e^{b(x-c)}\cdot[1+e^{d(x-e)}]}, \tag{1}$$

where, x is the displacement of the stereocilia; y is the receptor current; A is the maximal conductance; b and d are constants related to the slope or the sensitivity of the transduction; c and e are constants related to the resting position of the stereocilia. The reverse transduction process of the OHC has also been modeled by a first- or a second-order $f_B(x)$ relating transmembrane voltages to electromotile responses (Evans *et al.*, 1991; Santos-Sacchi, 1991). Such a model of the electro-mechanical transduction is based on the voltage-sensitive mechanical response mediated by the molecular motors residing in the body membrane of the OHC (Dallos *et al.*, 1991).

In vivo, cochlear MET transfer characteristics can be obtained by measuring the cochlear microphonic (CM), a far-field potential reflecting hair cell receptor currents. The MET function has been estimated from the CM in response to Gaussian noise (Chertoff *et al.*, 1996, 1997; Bian and Chertoff, 1998, 2001), to pure tones (Patuzzi and Moleirinho, 1998), and to tones modulated by a low-frequency tone (Nieder and Nieder, 1971). In this paper, we introduce a new method of estimating a function representing cochlear transduction from distortion product otoacoustic emissions (DPOAEs), which is noninvasive and may be useful in the evaluation of cochlear function in humans.

A. Theory for deriving a cochlear transducer curve from DPs

Assume that the transfer characteristics of cochlear transduction can be described by a nonlinear function $f_{NL}(x)$ which is n times differentiable at x_0; then, the value of $f_{NL}(x)$ can be approximated by an arbitrary input x using the Taylor's series expansion

[a]Electronic mail: lbian@kumc.edu

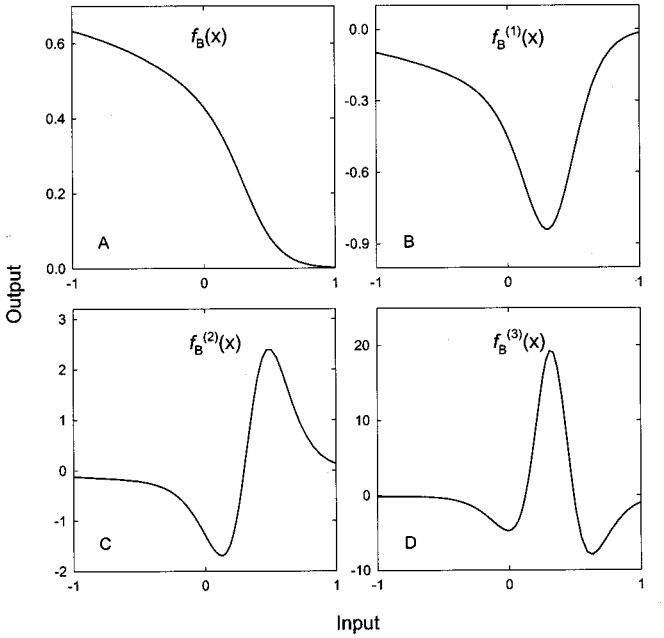


FIG. 1. A second-order Boltzmann function $[f_B(x)]$ representing cochlear MET input/output transfer characteristics (panel A). Panels B, C, and D are the first, second, and third derivatives of the $f_B(x)$, respectively. The input is normalized BM displacement, and output is receptor current in arbitrary units. A slope ratio (d/b) of 4.0 was adopted for the $f_B(x)$ (Kros *et al.*, 1992).

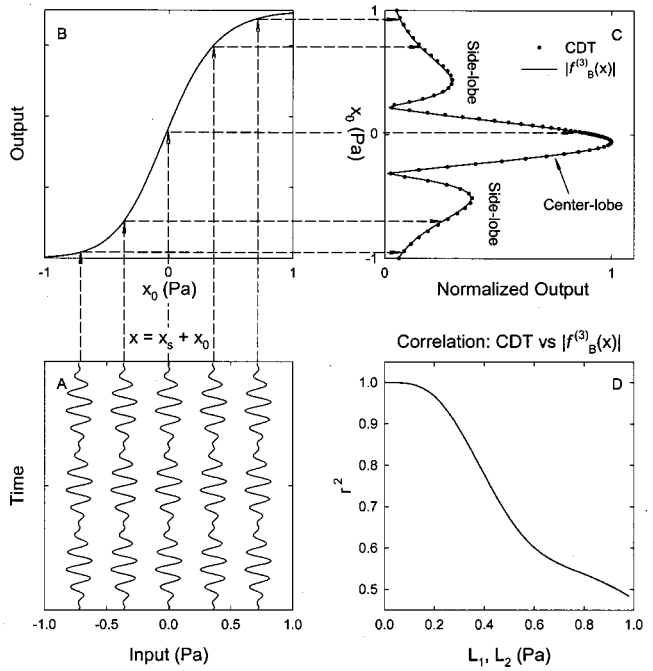


FIG. 2. Simulation: the effect of dc-shifting a two-tone signal on the CDT magnitudes. The two-tone signal (x_s) was dc-shifted to x_0 (panel A), then passed through a second-order $f_B(x)$ (panel B), and the CDT magnitudes were obtained by FFT (dots in panel C). The solid line in panel C represents the normalized absolute third derivative of the $f_B(x)$ or $|f_B^{(3)}(x)|$. The shape of the CDT magnitudes as a function of dc-shift consists of a center lobe and two smaller sidelobes. Panel D: the correlation (r^2) between the shape of CDT-dc function and the $|f_B^{(3)}(x)|$ improves asymptotically as the primary level ($L_1 = L_2$) reduces.

$$f_{NL}(x) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(x_0) \cdot (x - x_0)^n, \tag{2}$$

where $f^{(n)}(x_0)$ is the nth derivative of $f_{NL}(x)$ at the point of x_0, referred to as the operating point (OP). To estimate the whole function, a probe signal (x_s) could be systematically placed at different locations on the transduction curve [Figs. 2(A) and (B)] by adding a constant or dc shifting the x_s, i.e., $x = x_s + \text{dc}$. Physically, if the dc-shift changes the OP of the stereocilial bundle or the BM by a constant displacement during the stimulus, then it can be assumed that $\text{dc} = x_0$. Thus, Eq. (2) yields

$$f_{NL}(x) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(x_0) \cdot [(x_s + x_0) - x_0]^n. \tag{3}$$

Let $a_n = (1/n!) f^{(n)}(x_0)$, and when x_s is dc shifted in steps through the operating range (OR) of the cochlear transducer, then $f_{NL}(x)$ can be derived from a polynomial power series

$$f_{NL}(x_0) = \sum_{n=0}^{\infty} a_n x_s^n. \tag{4}$$

Stimulated by two pure tones, the inner ear generates various distortion products (DPs), of which the cubic difference tone (CDT, $2f_1 - f_2$) is the most prominent. Since DPs are produced from the nonlinearity in the cochlear transduction, they could be used to estimate the transducer function if the relation between DPs and the $f_{NL}(x)$ is known (Cheng, 1999). This relationship can be revealed by replacing the probe signal x_s in Eq. (4) with a two-tone stimulus, i.e., $L_1 \cos\theta_1 + L_2 \cos\theta_2$, and applying the binomial theorem and the trigonometry formula for $\cos^n\theta$ (for details, cf. Appendix A). As a result, the magnitude of the CDT (M_{2f1-f2}) can be expanded into a more explicit form (Engebretson and Eldredge, 1968; Lukashkin and Russell, 1998)

$$M_{2f_1-f_2} = L_1^2 L_2 \left| \left[\frac{3}{4} \frac{f^{(3)}(x_0)}{3!} + \frac{5}{8} \frac{f^{(5)}(x_0)}{5!} (2L_1^2 + 3L_2^2) + \frac{105}{64} \frac{f^{(7)}(x_0)}{7!} (L_1^4 + 4L_1^2 L_2^2 + 2L_2^4) + \cdots \right] \right|. \tag{5}$$

Therefore, M_{2f1-f2} depends on the primary levels and the absolute sum of the weighted third and higher odd-order derivatives of the $f_{NL}(x)$ evaluated at the OP (x_0). When the primary levels are sufficiently small, i.e., L_1, $L_2 \ll 1$, and as L_1, $L_2 \to 0$, the terms higher than the third-order in Eq. (5) become very small and negligible (Appendix A). Thus, the M_{2f1-f2} can be approximated as:

$$M_{2f_1-f_2} \approx \tfrac{1}{8} L_1^2 L_2 \cdot |f^{(3)}(x_0)|. \tag{6}$$

If the primary levels are held constant, M_{2f1-f2} is essentially proportional to the absolute value of the third derivative of the $f_{NL}(x)$ evaluated at different OPs, or $|f_{NL}^{(3)}(x_0)|$

$$M_{2f_1-f_2} \propto |f_{NL}^{(3)}(x_0)|. \tag{7}$$

Therefore, this makes it possible to obtain the third derivative of the cochlear transducer function by measuring CDT

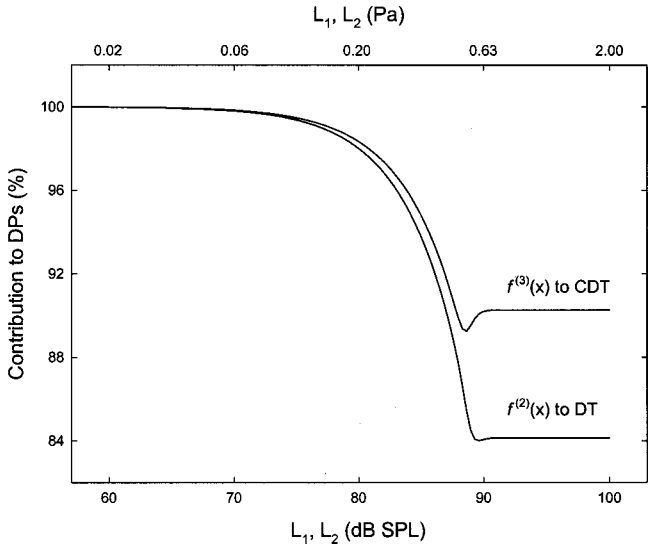


FIG. 3. Simulation: the contribution of the third derivative to the magnitude of the CDT ($2f_1-f_2$) at different primary levels ($L_1=L_2$). Also shown is the contribution of the second derivative to the amplitude of the DT (f_2-f_1).

magnitudes at various OPs across the OR of the cochlear transducer. Integrating the CDT-OP function, one could find the original $f_{NL}(x)$.

B. Simulations

Two computer simulations were conducted to numerically test the validity of the theory. The first simulation was to confirm that the third derivative contributes most to the M_{2f1-f2} from the mathematical expression of the CDT [Eq. (5)]. The second simulation was to determine the correlation between the CDT-OP function and the $|f_{NL}^{(3)}(x)|$, and examine this correlation as a function of the primary levels.

In the first simulation, the contribution (weight) of the third-order term to M_{2f1-f2} relative to a total contribution of 20 odd-order terms (including the third-order term) were compared for different primary levels ($L_1=L_2$) using the general form of the CDT [Eq. (A8)]. The primary levels ranged from 0 to 2 in linear unit, i.e., Pa. The results (Fig. 3) demonstrated that the third derivative accounts for almost 100% of the M_{2f1-f2} when the primary levels are below 0.06 Pa (70 dB SPL). Similarly, the even-order terms in the power series generate the difference tone (DT, f_2-f_1). The simulation showed that if the primary levels are below 70 dB SPL (0.06 Pa), the second derivative is responsible for almost 100% of the magnitude of DT (M_{f2-f1}) computed from a total of 20 even-order terms. Therefore, at lower primary levels, M_{2f1-f2} is dominated by the third-order term of the power series which approximates the $f_{NL}(x)$ [Eq. (4)], and M_{f2-f1} is primarily generated by the second-order term.

In the second simulation (Fig. 2), a two-tone signal (x_s) was dc shifted in both positive and negative directions in small steps (x_0) through the OR of a second-order $f_B(x)$ [Eq. (1)] representing a cochlear transducer function. It was assumed that the cochlear transducer operates within a range of $\pm$ 1 Pa (Dallos, 1986; Russell *et al.*, 1986). The output at each step was submitted to a fast Fourier transform (FFT) to obtain the CDT magnitude. As shown in Fig. 2(C), the CDT magnitudes varying as a function of the dc shift demonstrate a shape similar to that of the absolute value of the third derivative of the $f_B(x)$, denoted as $|f_B^{(3)}(x)|$. This shape contains a ''center lobe'' and two smaller ''sidelobes.'' When the primary levels ($L_1=L_2$) were reduced from 1 to 0 Pa, the correlation coefficient (r^2) between the CDT magnitudes as a function of dc-shift and the $|f_B^{(3)}(x)|$ improved asymptotically [Fig. 2(D)]. The r^2 reaches almost 1.0 when the primary levels are lower than 0.1 Pa (about 74 dB SPL). In other words, for low primary levels, the magnitude of CDT is proportional to the $|f_B^{(3)}(x)|$ [Eq. (7)].

C. Application in physiological experiments

Practically, presenting a low-frequency bias tone and a probe signal simultaneously could place the probe signal at different locations on the transducer curve. According to the mean-value theorem of calculus, if the probe signal is small enough, one can fairly accurately estimate the function. This principle was first used by Nieder and Nieder (1971) in determining the cochlear MET function from measuring the CM in response to a high-frequency tone which was modulated by a low-frequency bias tone. They found that the low-frequency modulated CM showed an envelope resembling the shape of the first derivative of an $f_{NL}(x)$. Numerically integrating the modulation envelope as a function of the bias level, they constructed a cochlear MET curve. Similarly, applying this idea and our theory to DPOAEs, one could obtain the shape of the absolute third derivative of an $f_{NL}(x)$ from measuring the low-frequency biased CDT magnitudes. Integrating the CDT-bias function three times or curve fitting this function by an equation of the $|f_{NL}^{(3)}(x)|$, the original $f_{NL}(x)$ could be found. This $f_{NL}(x)$ should reflect the transfer characteristics of the cochlear transduction, since the DPOAEs are produced from the nonlinearity in both the forward (Yates and Kirk, 1998) and reverse transduction processes (Santos-Sacchi, 1993). Therefore, we designed and conducted a physiological experiment on animals to test the adequacy of the theory.

II. METHODS

A. Experimental procedures

Nine Mongolian gerbils (*Meriones unguiculatus*), weighing between 47–63 g, were used as experimental subjects. The ''Animal Welfare Act'' and the ''Guide for the Care and Use of Laboratory Animals'' of the National Institute of Health (NIH) were followed in the experiments. The animals were anesthetized with Ketamine (100 mg/kg, i.p.) and maintained with a half of the initial dose given (i.m.) every hour. Rectal temperature was monitored and maintained at 37 °C with a heating pad (Harvard). The right pinna was removed and a speculum was sealed to the bony ear canal. A calibrated probe microphone (Etymotic Research, ER-10B) was inserted into the speculum about 5 mm from the umbo of the tympanic membrane to record the signal. The acoustic stimuli, two primary tones and a bias tone, were generated from an array processor (Tucker-Davis Tech., TDT AP2) and downloaded to a three-channel D/A converter

(TDT DA3-4) which was routed to three attenuators (TDT PA4). The two tones were sent to a headphone buffer (TDT HB6) connecting to two earphones (ER-3A) that were coupled to the ports on the ER-10B. The bias tone was passed to an amplifier (Optimus SA-155) leading to a subwoofer (Boston Acoustics) that was connected with a silicon tube to a port near the tip of the speculum.

Prior to each recording session, the CDT magnitude was obtained on a spectral analyzer (Hewlett-Packard, 3561A) with the primary levels set at 80 dB SPL and no bias tone. A CDT magnitude greater than 40 dB SPL was considered normal. This CDT magnitude was monitored and maintained at 40 dB SPL with variations less than 5 dB throughout the experiment. Each recording session included recordings of the ear canal acoustics while presenting a two-tone signal and a bias tone with various amplitudes. Specifically, with the primary level set and held constant, the amplitude of the bias tone was attenuated systematically from 20 to 0 Pa in 21 steps, with more steps for lower bias levels. There was a recording session for each of the five different primary levels. Then, the procedure was repeated with an interval of about 90 min.

B. Signal processing and data analysis

Two pairs of primary tones were used in the experiment: 3968 and 5120 Hz ($f_2/f_1=1.29$), 5888 and 7552 Hz ($f_2/f_1=1.28$). Such a relatively large f_2/f_1 ratio was selected to minimize the contribution from the CDT frequency place to the generation of the CDT (Knight and Kemp, 2000). Since DPOAEs are generated from a relatively localized region tuned to the geometric mean of the two primary frequencies or more closer to f_2 (Brown and Kemp, 1984; Martin *et al.*, 1987; He and Schmiedt, 1997), especially at low to moderate signal levels (Avan *et al.*, 1998), the $f_{NL}(x)$ derived from the CDT should reflect the cochlear transfer characteristics within the overlapped frequency region. Primary levels ($L_1=L_2$) ranging from 50 to 70 dB SPL in 5-dB increments were adopted in the experiment. Since the f_2/f_1 ratio is fixed, at a given primary level, the possible two-tone suppression should be steady and also can be attributed to the nonlinear transducer function (Geisler and Nuttall, 1997). The frequency of the bias tone (f_0) was 25 Hz with peak amplitudes ranging from 20 Pa (120 dB SPL) to 0 Pa. The signal level of the bias tone was calibrated in the speculum at 120 dB SPL with a sound-level meter (Quest 1800). Figure 4 shows an example of the acoustic signal recorded from the ear canal. The bias tone was 250 ms in duration, with the tail portion (about 40 ms) being flat. This made it possible to check the reliability of the recordings by verifying that the CDT magnitudes at the tails were consistent throughout the experiment.

The acoustic signal recorded by the ER-10B was low-pass filtered at 16 kHz, amplified 10 times (Stewart VBF 10M), and digitized at 65 536 Hz (TDT AD2). Depending on the primary levels, the signal was averaged 30 to 150 times with more averages at lower primary levels. The data was analyzed off-line using MATLAB (6.0 MathWorks). The signal was high-pass filtered at 1000 Hz to eliminate the bias tone. The segments, each containing 512 points, corresponding to the four peaks and troughs with consistent magnitudes in the original signal (Fig. 4), were Hanning-windowed and submitted to an FFT to extract the CDT magnitudes. Furthermore, the original acoustic signal was also low-pass filtered at 100 Hz to obtain the bias tone. The positive and negative peak amplitudes of the bias tone were obtained and considered as the bias levels. The CDT magnitudes and their corresponding bias levels were paired as a data set. Ideally, the transducer function should be derived from integrating the CDT-bias function three times. However, each step in the multiple integration requires a constant that is unavailable from the data, therefore, we adopted a curve-fitting method to derive the transducer function. A function representing the $|f_B^{(3)}(x)|$ [Eq. (B10)] was fit to the data set to obtain the parameters of the second-order $f_B(x)$. To assess the effects of the experiment conditions on the $f_B(x)$ resulting from the curve fitting, the parameters of the $f_B(x)$ were submitted to a three-way (time, frequency, level) analysis of variance (STATISTICA 5.0, StatSoft). A p value of less than 0.05 was considered statistically significant. The test–retest reliability of the CDT data was evaluated by the correlation between the CDT magnitudes measured at time 1 and time 2.

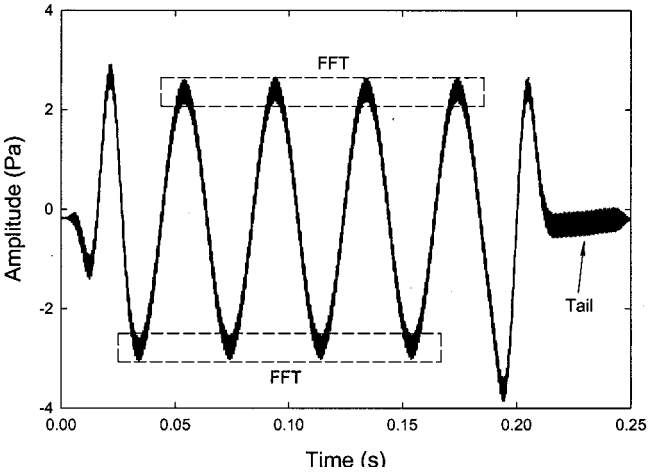


FIG. 4. An example of the acoustic signal recorded from the ear canal. The stimulus contains a bias tone ($f_0=25$ Hz) and two primaries ($f_1=3968$ Hz, $f_2=5120$ Hz). The peak amplitude of the bias tone (L_0) is 102 dB SPL (2.5 Pa) and the primary level ($L_1=L_2$) is 60 dB SPL (0.02 Pa). The two rectangles indicate the four peaks and troughs where the CDT magnitudes were obtained by FFT. The tail portion of the signal, where the bias level is zero, serves as a control for reliability.

III. RESULTS

A. CDT magnitudes

In the time domain, the CDT can be extracted from the acoustic signal by a narrow bandpass filter centered at the CDT frequency. Figure 5 shows the bandpassed (equiripple FIR filter with 200 Hz bandwidth) acoustic signal from an animal who received a 55-dB SPL two-tone signal ($f_1=3968$ and $f_2=5120$ Hz) biased by a 25-Hz tone with 120-dB SPL peak amplitude. The CDT magnitude varies consistently and systematically as the amplitude of the bias tone varies. The CDT magnitude is highest when the bias tone crosses the x axis and at the tail portion where no bias tone is present. As the bias tone amplitude increases, the

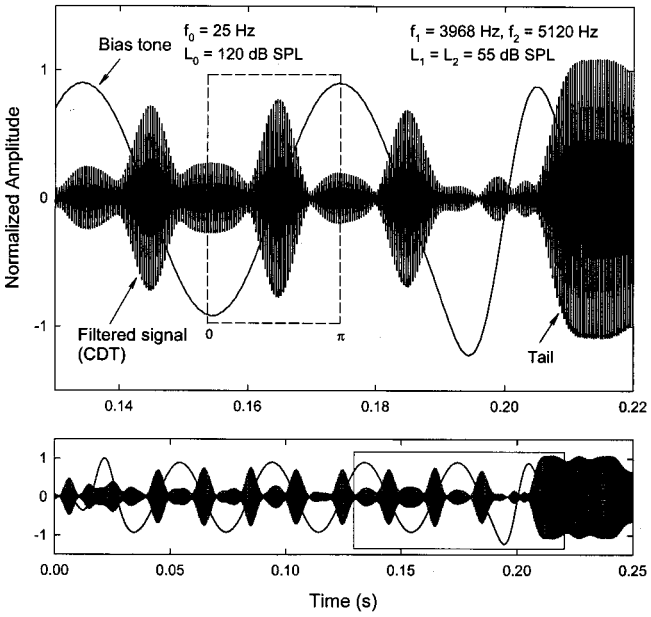


FIG. 5. A time-domain representation of the CDT. The signal was obtained by passing the ear canal acoustic signal through an equiripple FIR filter (bandwidth=200 Hz) centered at the CDT frequency. The upper panel shows a magnified view of a section (rectangle of solid line) from the full signal in the lower panel. The rectangle of dashed line in the upper panel demonstrates the CDT modulation pattern within 1/2 cycle of the bias tone. L_0 represents bias tone level.

CDT is suppressed. After reaching a minimum, the CDT magnitude grows slightly when the bias tone level rises to its peak amplitude.

The effect of the bias tone on CDT can be quantified by obtaining a CDT-bias function, which is a series of CDT magnitudes obtained at the peak and trough regions in the signals with different bias levels. The CDT-bias functions obtained from all the animals for different primary frequencies and levels are shown in Figs. 6 and 7. Since the signal recorded in the experiment contained four consistent peaks and troughs (Fig. 4), four CDT magnitudes were obtained from a single bias level. The differences between the four values (vertical bars in Figs. 6 and 7) can serve as within-stimulus variability. The CDT-bias functions demonstrate a similar pattern across all the frequencies and levels of the primaries. The CDT magnitude is highest when the bias level is about zero, and reduced when the bias level increases in both positive and negative directions, forming a center lobe. The CDT magnitude reaches a minimum around bias levels of ± 10 Pa, and then increases as the bias level continues to grow forming two smaller sidelobes. This CDT modulation pattern is similar to the simulation and the shape of the $|f_B^{(3)}(x)|$ [Fig. 2(C)]. The test–retest reliability of the CDT magnitude is generally better at lower signal levels. The r^2 between the CDT magnitudes measured at trial 1 and trial 2 averaged across all the animals are greater than 0.73 for primary levels between 50 and 60 dB SPL with the highest r^2 at 55 dB SPL (right panels in Figs. 6 and 7). For signal levels above 65 dB SPL, the average r^2 ranged from 0.48 to 0.63 accompanied by higher between-subject variability.

From the CDT magnitudes averaged across all the animals for different primary levels and frequencies (Fig. 8), the typical shape of the $|f_B^{(3)}(x)|$ presents at 55-, 60-, and 65-dB SPL primary levels. The amplitude and the width of the center lobe increase as the primary level increases. At 50 dB SPL, the center lobe is narrower and sidelobes are barely noticeable. At 70 dB SPL, the center lobe expands to almost the whole range of the bias levels ($\pm$ 20 Pa). Comparing the two pairs of primary frequencies, the magnitude of the center lobe is higher for the higher primary frequency.

B. Deriving a Boltzmann function from the CDT data

For each animal, the CDT-bias function at every primary level was fit with an equation that represents the $|f_B^{(3)}(x)|$ [Eq. (B10)]. Curve fitting was conducted in MATLAB using a function minimization routine to reduce the sum of the squared deviations between the data and the fit. The set of parameters chosen as the fitting result was the set that had highest r^2, least error, and produced a sigmoidal curve with a positive slope. Since the CDT magnitudes obtained at time 1 and time 2 were well correlated for most primary levels, the mean and standard deviation (s.d.) of the parameters of the $f_B(x)$, as well as the r^2 of the fits, were averaged across trial 1 and trial 2 (Table I). The highest r^2 of the fits is at 55 dB SPL for both primary frequencies (0.93 and 0.88 for f_1 of 3968 and 5888 Hz, respectively) followed by 60, 50, and 70 dB SPL. The lowest r^2, 0.63 is present at the 65 dB SPL condition.

The fits of the CDT magnitudes averaged across all the subjects are shown in Fig. 9 (top panels). Compared to the mean CDT data in Fig. 8, these averaged fits generally represent the mean data. Even though the parameters obtained from the curve fitting show considerable variability, there are still consistencies that can be observed. For the slope parameters b and d,b is negative and the absolute value of b is always larger than d. The mean absolute value of the b/d ratio is 2.1. This indicates that the sigmoidal curve of the $f_B(x)$ is asymmetrical. For parameters c and e,c is negative and more consistent, but e is highly variable. Statistical analysis did not reveal any significant frequency and level effects or frequency$\times$level interaction for the parameters besides A. Parameter A has a significant level effect with greater values for higher signal levels. The bottom panels of Fig. 9 show the averaged sigmoidal curves plotted from the parameters of the $f_B(x)$ fit to each animal. They all demonstrate a similar sigmoidal shape, but with a progressive dc-shift as the primary level increases. There is an increase in the range of the output or dynamic range (DR) when the signal level rises. Comparing across frequency, the DR is smaller for higher primary frequencies.

IV. DISCUSSION

A. Modulation of DPOAE by bias tone

The results of the experiment demonstrate that the low-frequency bias tone causes a modulation of the DPOAE magnitude, specifically $2f_1-f_2$. This is revealed by not only the narrow-bandpassed signal centered at the CDT frequency (Fig. 5), but also the CDT magnitudes retrieved from the acoustic signals when the two-tone input was modulated by a series of bias tones with different amplitudes (Figs. 6 and 7).

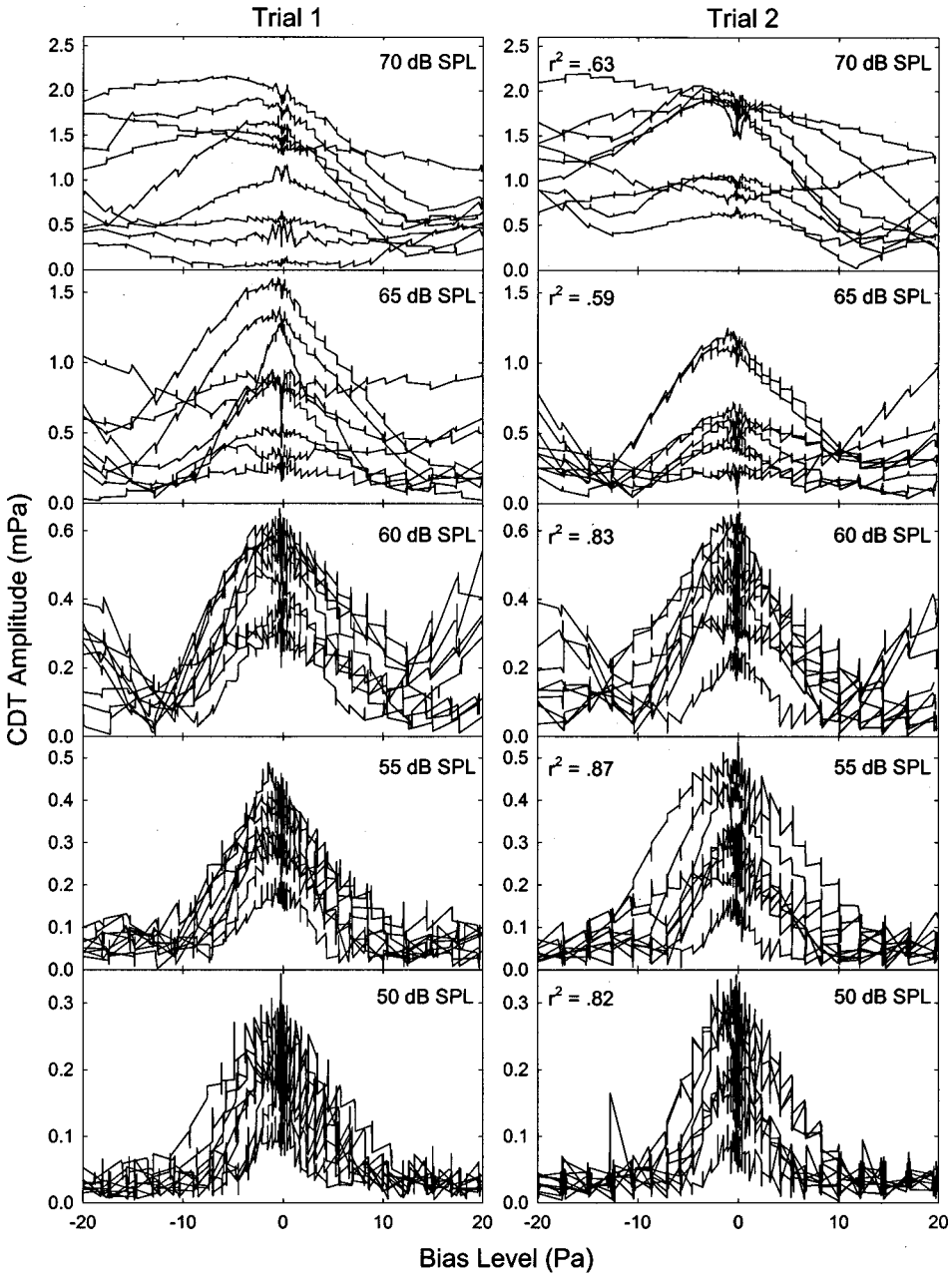


FIG. 6. The CDT magnitudes as a function of the bias levels obtained for trial 1 (left panels) and for trial 2 (right panels) from nine animals (1): $f_1 = 3968$ and $f_2 = 5120$ Hz. From the top panels to the bottom, the primary levels ($L_1 = L_2$) are 70, 65, 60, 55, and 50 dB SPL, respectively. The short vertical lines represent the within-stimulus variability of the CDT magnitudes obtained from the four peaks and troughs. Also indicated in the right panels are the average r^2 between the CDT data obtained at trial 1 and trial 2.

The low-frequency modulation of the CDT magnitudes in the present study is consistent with the results of Frank and Kössl (1996, 1997) in gerbils, and Scholz *et al.* (1999) in humans. Such modulation of the CDT also has been found in electrically evoked otoacoustic emission (EEOAE) in gerbils (Kirk and Yates, 1998). The present study shows that the maximal CDT occurred at zero crossings of the bias tone magnitude, which agrees with the findings of Frank and Kössl (1996, 1997).

The reason for the modulation of the DPOAEs by the bias tone could partly be attributed to the shifting of the resting position of the BM. When the BM is biased, whether it is towards the scala vestibuli (SV) or scala tympani (ST), the ac receptor potentials of the OHCs are suppressed (Russell and Kössl, 1992a; Cheatham and Dallos, 1994). This, in turn, could reduce the voltage-sensitive motile response of the OHCs (Santos-Sacchi, 1989) and lead to reductions in the DPOAEs. Another reason could be that the bias tone drives the OHCs into an asymptotic region on the nonlinear MET curve [Fig. 1(A)], where the gain of the MET is less optimal. Therefore, the displacement of the stereocilial bundle could produce smaller voltage change, thus leading to smaller motile response and DPOAEs.

Interestingly, different regions on the transducer curve could be responsible for the differences in the amplitudes of various DPs. For example, when the two-tone signal is placed in the middle part of the transducer curve, where the curve contains mainly odd symmetry, it gives rise to mostly CDT and little DT. Figure 10 shows the effect of the bias tone on the DT from five animals. The DT magnitude demonstrates a notch at the center or 0 Pa bias level where the CDT magnitude shows a center lobe (Figs. 6, 7). If the two-tone signal is biased by about ± 10 Pa, the DT magnitude is maximal, while the CDT magnitude shows two notches. Comparing to the transducer curves in Fig. 9 (bottom panels), the function starts to saturate beyond these regions

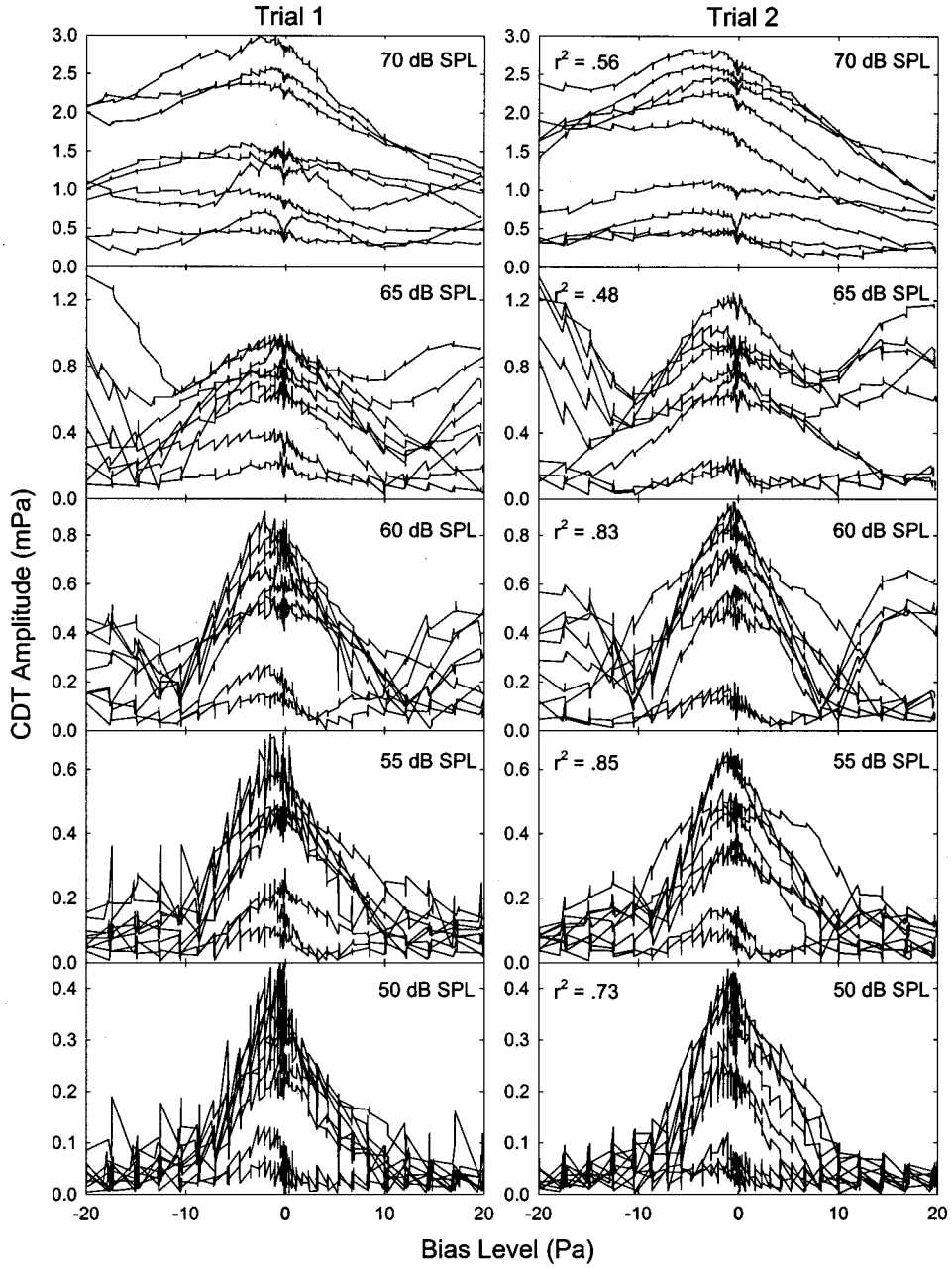


FIG. 7. The CDT magnitudes as a function of the bias levels obtained for trial 1 (left panels) and for trial 2 (right panels) from nine animals (2): $f_1 = 5888$ and $f_2 = 7552$ Hz. The representations of the labels in each panel are the same as Fig. 6.

(±10 Pa). As proposed by van Emst *et al.* (1997), the saturation nonlinearity consists of even symmetry, thus producing a greater DT. They observed an increase in the even-order DPs while increasing the dc component of the cochlear potential by current injection or perfusion of furosemide. An increase in the DT was also found following a brief exposure of a 200-Hz tone (Kirk and Patuzzi, 1997). They attributed the increase of the DT to the shift of the OP on the transducer function. Furthermore, opposite changes in the even- and odd-order DPs also have been found in the OHC gating currents when dual-sinusoid voltage steps were applied to the isolated cells (Takahashi and Santos-Sacchi, 1999). Therefore, the generation of the DPs depends upon not only the shape of the nonlinear transduction curve, but also the location of the OP on the curve. It is noteworthy to point out that the magnitude of DT as a function of bias level resembles the shape of the $|f_{NL}^{(2)}(x)|$ as predicted in the theory. However, the DT data were not included in the analysis due to high intersubject variability.

B. Effects of signal levels on CDT

1. Bias tone level

The greatest suppression of the CDT occurred at a bias level of about ±10 Pa, which corresponds to about 114 dB SPL. Such a bias level is comparable to the signal levels adopted in the biasing experiments by others, e.g., 112 (Kirk and Yates, 1998) and 115 dB SPL (Frank and Kössl, 1996, 1997; Scholz *et al.*, 1999). Their CDT data showed a modulation pattern of the CDT magnitude that consisted of a larger peak and a smaller peak within one cycle of the bias tone, similar to the finding in the present study (Figs. 5, 6, and 7). Extending the bias level to 120 dB SPL in the present experiment revealed the sidelobes in the CDT-bias function.

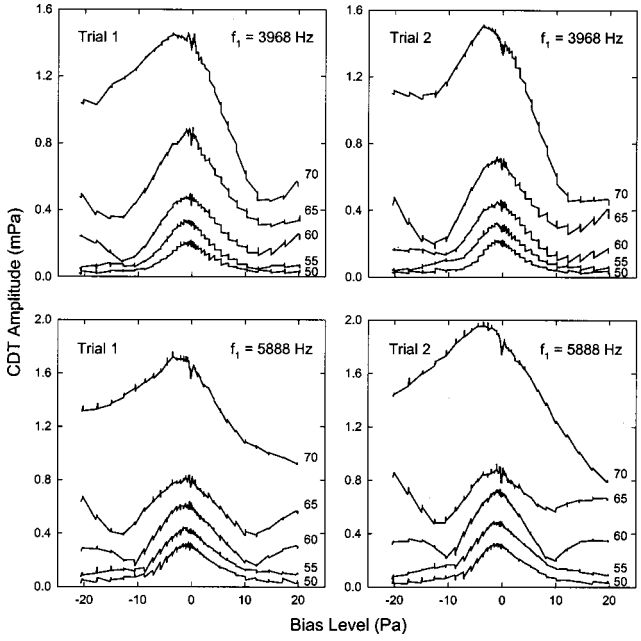


FIG. 8. The averaged CDT magnitudes as a function of bias levels. The data obtained for trial 1 and for trial 2 are shown in the left and right panels, respectively. Top and bottom panels represent different primary frequencies. The labels next to the curves indicate the primary levels ($L_1=L_2$).

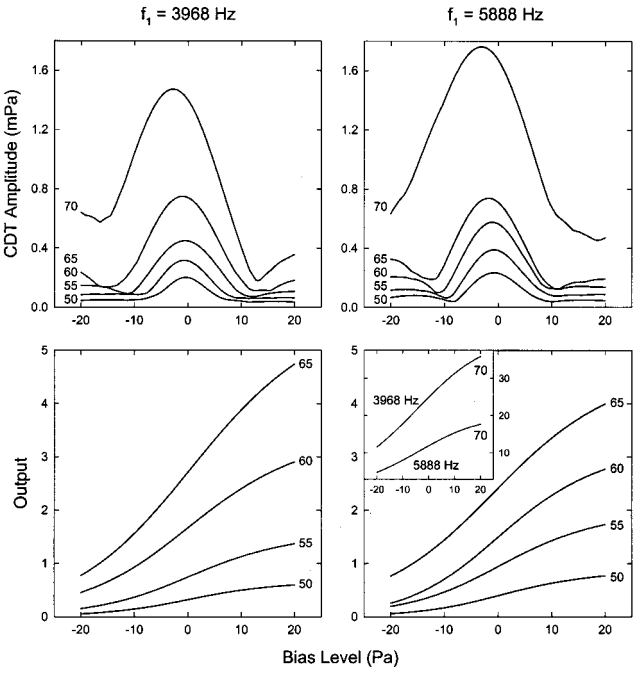


FIG. 9. The averaged fits of $|f_B^{(3)}(x)|$ to the CDT-bias functions (upper panels) and the averaged second-order $f_B(x)$ (lower panels) across all the animals. Left and right panels represent the different primary frequencies. The labels in each panel indicate the primary levels ($L_1=L_2$). The insert shows the $f_B(x)$ obtained at 70-dB SPL primary levels.

The bias levels were adopted as a representation of the actual input to the derived transducer function (lower panels of Fig. 9). The mechanical tuning curve of the BM shows a 40–60-dB difference between the tip and the tail regions (e.g., Ruggero *et al.*, 1997; Nilsen and Russell, 2000). The actual displacement of the BM by the bias tone at the f_1 and f_2 overlap region could be 40 to 60 dB lower than the BM location centered at 25 Hz. Therefore, the maximal effective bias level at the BM location where the DPOAEs are generated could be 60 to 80 dB SPL, which is consistent with the saturation level of the OHCs receptor potentials (Dallos, 1986; Russell and Kössl, 1992b). Thus, the range of ±120-dB SPL bias level which corresponds to about ±80-dB SPL effective input range is adequate to represent the OR of a cochlear transducer function at the DPOAE site.

2. Primary level

The mean CDT magnitude in dB SPL from Fig. 8 averaged across the two trials can be summarized in a 3D surface plot shown in Fig. 11 for the two pairs of primary frequencies. As the primary level grows, three trends can be observed: (1) the magnitudes of the center- and sidelobes increase; (2) the modulation depth, which is the height of the center lobe relative to the maximal suppressions on either direction of the bias level, decreases; and (3) the width of the center lobe increases. Such interaction between the primary level and the bias level on the CDT magnitude can produce different growth patterns depending upon where the OP is set. This could account for the various patterns of the CDT

TABLE I. Curve-fitting results: the mean and s.d. of the parameters of a second-order Boltzmann function whose third derivative was fit to the CDT data and the correlation coefficients (r^2) of the fits for different primary frequencies and levels. The means reflect an average of the two trials.

f_1 (Hz)	L_1, L_2[a] (dB SPL)	A	b	c	d	e	r^2
3968	50	0.70±0.80	−0.16±0.06	−0.90±2.0	0.09±0.15	11.3±7.7	0.88±0.08
	55	1.7±2.0	−0.15±0.09	−5.9±9.9	0.05±0.10	8.5±7.8	0.93±0.03
	60	3.9±3.5	−0.15±0.08	−5.6±7.3	0.07±0.07	−2.4±14.7	0.90±0.06
	65	7.3±9.2	−0.10±0.05	−5.7±12.6	0.00±0.11	3.0±13.8	0.77±0.32
	70	24.0±22.0	−0.13±0.08	−6.4±6.2	0.07±0.08	−7.1±16.6	0.81±0.15
5888	50	1.0±0.9	−0.15±0.07	−1.8±8.3	0.07±0.11	5.2±7.5	0.75±0.31
	55	2.4±2.6	−0.15±0.04	−0.90±4.2	0.09±0.07	9.0±10.3	0.88±0.14
	60	3.7±3.8	−0.18±0.08	−6.2±4.4	0.09±0.10	−9.2±13.6	0.86±0.13
	65	5.7±8.6	−0.15±0.10	−4.5±9.6	0.05±0.13	−1.7±16.9	0.63±0.31
	70	46.1±37.8	−0.09±0.05	−13.7±14.3	0.05±0.07	−6.2±42.0	0.81±0.14

Mean±s.d.
[a]$L_1=L_2$.

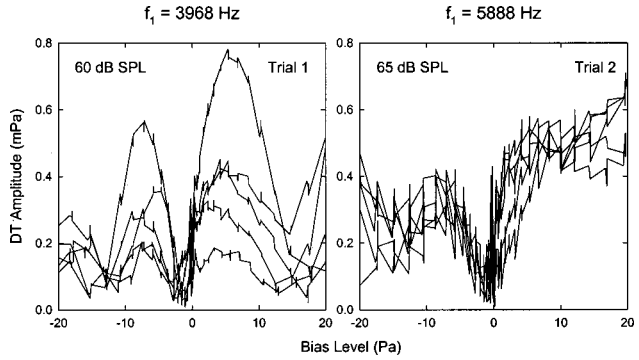


FIG. 10. Samples of the DT magnitudes as a function of the bias levels from five animals. The DT magnitudes were obtained at trial 1 for f_1 of 3968 Hz (left panel); at trial 2 for f_1 of 5888 Hz (right panel). The signal level in each panel represents the primary levels ($L_1 = L_2$).

growth function found by many investigators in both humans (Lonsbury-Martin *et al.*, 1990; Nelson and Kimberley, 1992; Popelka *et al.*, 1993; Abdala, 2000) and animals (Brown, 1987; Lonsbury-Martin *et al.*, 1987).

The modulation depth of the CDT is largest (20 dB) at the lowest primary level (Fig. 11) which is consistent with the findings of others. Scholz *et al.* (1999) reported a 17.7-dB suppression of the CDT by a 32.8-Hz tone ($L_1 = 50$, $L_2 = 25$ dB HL). Frank and Kössl (1996) showed a nearly 20-dB reduction in CDT magnitude when the primaries ($L_1 = 48$, $L_2 = 38$ dB SPL) were biased by a 5-Hz tone. The modulation depth could reflect the change in the slope or gain of the transducer curve when the OP is shifted from about 0 to ± 10 Pa bias level. If biasing in only one direction is considered, as the slope of the transduction curve reduces, the CDT magnitude shows level-dependent changes (Figs. 8, 9, and 11) that resemble the experimental results when the cochlear gain was reduced. For example, after injection of furosemide, the CDT magnitude and phase as a function of time showed nonmonotonic changes marked by a deep notch and a phase reversal at lower primary levels (Mills *et al.*, 1993; Mills and Rubel, 1994). Similar results were also observed when cochlear blood flow was blocked (Mom *et al.*, 2001). Such changes in CDT magnitude and phase could be explained by the OP shift along the cochlear transducer curve. Furosemide can shift the OP of the cochlear transducer function (Bian and Chertoff, 2001), and this OP shift could increase with time as the drug accumulates in the cochlea. Similarly, ischemia could shift the OP by reducing in the endocochlear potential (Konishi *et al.*, 1961). The ''active'' CDT as proposed by Mills (1997), corresponding to the center lobe, comes from the middle portion of the sigmoidal function with highest gain (slope), whereas the ''passive'' CDT, represented by the sidelobes, is produced from the saturation region on the transduction curve where the gain (slope) is smaller. Moreover, the CDTs generated from these two regions on the transducer curve always have a 180° phase difference, since the underlying directions of the center- and the sidelobes are opposite [Fig. 1(D)]. Thus, the CDT magnitude reflects the gain (slope) of the cochlear transducer function. Furthermore, the effect of primary levels on the CDT modulation depth in the present study and the studies discussed above indicates that the optimal gain of the transducer is higher for lower signal levels (Bian and Chertoff, 1998) or the existence of positive feedback (Lukashkin and Russell, 1999).

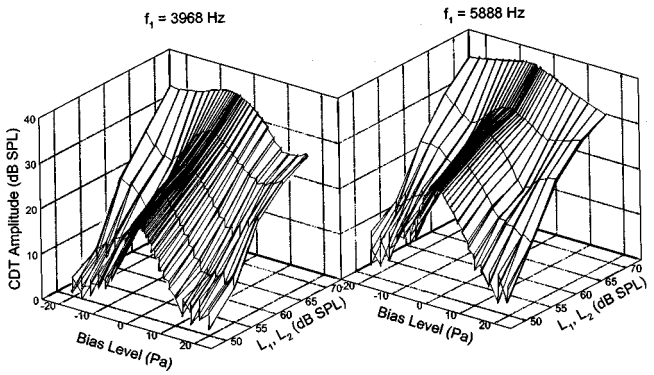


FIG. 11. 3D surface plot of the mean CDT-bias function for f_1 at 3968 Hz (left panel) and f_1 at 5888 Hz (right panel). The CDT magnitudes are averaged across the two trials for all the animals. The surface plots demonstrate various growth patterns depending on where the OP (bias level) is set.

There are noticeable changes in the symmetry of the CDT-bias function as the primary levels increase. The peak of the center lobe shifts slightly to the negative bias levels from zero Pa (Fig. 8). In addition, at higher primary levels, the depths of the two modulation notches become uneven, with the deeper notch at the positive bias level. Occasionally, the sidelobe is taller on the negative side of the bias level, a direction that is consistent with a biasing of the BM towards SV (Cheatham and Dallos, 1994). This may be comparable to the results of Kirk and Yates (1998) who demonstrated an enhancement of the EEOAE magnitude when the BM was biased toward SV. Shifting the cochlear partition towards SV could depolarize the OHC, thus increase the motile response of the OHCs (Santos-Sacchi, 1989). At lower primary levels, the CDT-bias function is more symmetrical. This is consistent with the finding *in vivo* that the OHC electrical response is symmetrical at low signal levels (Russell *et al.*, 1986). The symmetrical transduction characteristic indicates that the cochlear transducer operates freely with its optimal gain at low signal levels.

C. Transduction characteristics

At moderate to low signal levels, the nonlinearity in MET is predominately responsible for the mechanical response of the OHCs (Santos-Sacchi, 1993). Thus, the $f_{NL}(x)$ derived from the CDT-bias function in the present study resembles the transfer characteristics of MET. The transducer functions obtained from the CDT data show an increased DR and a reduced slope at high signal levels. This is consistent with our previous findings (Bian and Chertoff, 1998) and the results of others (Patuzzi and Moleirinho, 1998), who showed a system of sigmoidal curves of MET with varying DR and slopes as a function of the input level. Another major level effect is the greater dc shift of the whole transduction curve at higher primary levels. This is compatible with the observations that the dc receptor potential of the OHC increases with signal level (Russell *et al.*, 1986). An increase in the transmembrane potential at higher signal levels in turn

could produce large dc components in the OHC motile response (Santos-Sacchi, 1989; Evans *et al.*, 1991).

Comparing the two pairs of primary frequencies, the transduction curves for lower primary frequency generally have a greater DR than higher frequency primaries. This is in agreement with the shapes of the transduction curves obtained from the CM evoked by broadband noise (Bian and Chertoff, 2001) and OHC receptor potential at different cochlear locations (Dallos, 1986). These studies showed that MET transfer functions in the low-frequency region have a greater output range than in the high-frequency region. The $f_{NL}(x)$ obtained from CDT data in the present study is more symmetrical (average $b/d=2.1$) than the MET functions observed from *in vitro* studies (e.g., Kros *et al.*, 1992). It is possible that the transduction in intact cochlea is more symmetrical, and perhaps the transducer function obtained from DPOAEs contains both forward and reverse transduction processes.

D. Technical issues

When we fit the $|f_B^{(3)}(x)|$ to the CDT data, the resulting parameters were not unique, i.e., different parameters could produce an equally good fit to the data, even when the r^2 was greater than 0.9. One solution to this problem was to normalize the CDT function by taking the ratio of the CDT magnitudes at the biasing peaks and at the tail portion of the stimulus where there is no biasing, and fit the data with a first-order $f_B(x)$ instead of the second-order $f_B(x)$. This method provided more consistent parameters, but the first-order $f_B(x)$ is symmetrical about its inflection point and cannot truly reflect the asymmetry of the cochlear transduction.

The noise floor of the spectral analysis in the present study was about 5 dB SPL. The theory suggests that the lowest primary level should be used to derive the $f_{NL}(x)$, but practically the CDT would fall into the noise floor, e.g., the sidelobes of the CDT-bias function are barely noticeable at 50-dB SPL primary level (Figs. 6 and 7). This could result in inaccurate estimates of the original transducer function. It is especially crucial for future experiments in the ears with cochlear hearing loss or in humans, where the CDT magnitude is smaller. One reason for the high noise floor could be that the number of data points at each bias peak or trough included in the FFT was limited to 512. Possible solutions to this problem are faster sampling rate, and lower bias frequency. More amplifications and averages also can improve the signal-to-noise ratio.

At 70 dB SPL primary level, there is considerable between-subject variability in the CDT magnitude data, even though the within-subject variability is small (Figs. 6 and 7). As indicated in Eq. (5), when the primary levels are relatively high, the CDT magnitudes may have contributions from the higher derivatives of the transducer function. In a similar derivation of the harmonic distortions for single tone from a power series, Weiss and Leong (1985) pointed out that for moderate to high signal levels the higher-order terms play a role in the generation of the DPs. Fahey *et al.* (2000) suggested that the coefficients of the higher terms remain small until the cochlear transducer is driven into saturation. Moreover, when the third derivative is zero (e.g., at ± 10-Pa bias level), the next higher-order term dominates the generation of CDT. These higher-order terms may be more susceptible to subtle changes in cochlear function, and the interaction of their phases are more complex, thus increasing the variability and reducing the repeatability of the CDT measure at higher primary levels (>65 dB SPL). This may also account for the inaccuracy for using just the $|f_B^{(3)}(x)|$ to estimate the transducer function at higher primary levels. From this experiment, we found that the optimal primary levels to obtain the cochlear transducer function are 55 and 60 dB SPL, where both the reliability and the correlation of CDT-bias function with the $|f_B^{(3)}(x)|$ are among the highest. In future research, including the higher-order terms in the analysis for DPs may help to extend the theory to higher primary levels and characterize the cochlear transducer function more accurately.

V. SUMMARY AND CONCLUSION

The results of this study showed that the magnitudes of DPOAEs were modulated by simultaneously presenting a low-frequency bias tone. For primary levels below 70 dB SPL, the modulation pattern of the CDT magnitude by the bias tone closely resembled the shape of the third derivative of a sigmoidal function. Curve fitting allowed the derivation of a second-order $f_B(x)$ that characterizes cochlear transduction. The results of the experiment were in good agreement with the simulation of passing dc-shifted two-tone signals through an $f_B(x)$ that represents cochlear transduction. The results of the experiment were also consistent with the theory that at low primary levels the CDT magnitude is proportional to the third derivative of a nonlinear function that governs the generation of DPOAEs. There was also a level effect on the shape of the CDT-bias function. The modulation depth was greater and the center lobe was narrower at lower primary levels. The sidelobes are most prominent at 55- and 60-dB SPL primary levels, which were found to be optimal to fit the $|f_B^{(3)}(x)|$. The derived $f_B(x)$ also showed level effects such as a dc shift and increase in the DR as primary level grows. The present study demonstrates that DPOAEs can be used to estimate the transfer characteristics of the cochlear transduction. The noninvasive nature of DPOAE measure could lead to the clinical application of this technique in the evaluation of cochlear function.

ACKNOWLEDGMENTS

The authors wish to thank David Lerner for his valuable comments on the manuscript. This study was supported in part by the National Institute on Deafness and other Communication Disorders of NIH, Grant No. 2R01 DC02117, and USPHS Grant No. HD02528 to the Smith Mental Retardation and Human Development Research Center at the University of Kansas Medical Center.

APPENDIX A: DERIVING CDT FROM A POLYNOMIAL POWER SERIES

From Eq. (4), given a power series $f(x)=\Sigma_{n=0}^{\infty} a_n x^n$, and $a_n=(1/n!)f^{(n)}(x_0)$, let x be a two-tone signal, i.e., x

$=x_1+x_2$, where $x_1=L_1\cos(2\pi f_1 t)$, $x_2=L_2\cos(2\pi f_2 t)$, then the output of the nonlinear system is defined as

$$y(x_1,x_2)=\sum_{n=0}^{\infty} a_n(x_1+x_2)^n. \tag{A1}$$

Since $(x_1+x_2)^n=\Sigma_{k=0}^{n}\binom{n}{k}x_1^{(n-k)}x_2^k$, and let $h=n-k$, then

$$y(x_1,x_2)=\sum_{n=0}^{\infty}\left[a_n\sum_{k=0}^{n}\binom{n}{k}x_1^h x_2^k\right], \tag{A2}$$

where $\binom{n}{k}$ are binomial coefficients. Substitute x_1 and x_2 with their sinusoid forms and let $\theta_1=2\pi f_1 t$, and $\theta_2=2\pi f_2 t$, then the output as a function of the two primary frequencies is

$$y(\theta_1,\theta_2)=\sum_{n=0}^{\infty}\left[a_n\sum_{k=0}^{n}\binom{n}{k}(L_1\cos\theta_1)^h(L_2\cos\theta_2)^k\right]. \tag{A3}$$

In terms of frequency composition, $\cos^n\theta=(1/2^{n-1})\Sigma_{i=0}^{m}\binom{n}{i}\cos(n-2i)\theta$, where $m=n/2-1$ or $(n-1)/2$, depending on whether n is even or odd. Substituting Eq. (A3) with this type of expression yields

$$y(\theta_1,\theta_2)=\sum_{n=0}^{\infty}\left\{a_n\sum_{k=0}^{n}\left[\binom{n}{k}\cdot\frac{L_1^h}{2^{h-1}}\sum_{i=0}^{p}\binom{h}{i}\cos(h-2i)\theta_1\cdot\frac{L_2^k}{2^{k-1}}\sum_{j=0}^{q}\binom{k}{j}\cos(k-2j)\theta_2\right]\right\}, \tag{A4}$$

where $k<n$; and

$$p=\begin{cases} h/2-1, & (h:\ \text{even})\\ (h-1)/2, & (h:\ \text{odd})\end{cases},\quad q=\begin{cases} k/2-1, & (k:\ \text{even})\\ (k-1)/2, & (k:\ \text{odd})\end{cases}.$$

Rearranging Eq. (A4), the output spectral components can be given as

$$y(\theta_1,\theta_2)=\sum_{n=0}^{\infty}\left\{\frac{a_n}{2^{n-2}}\sum_{k=0}^{n}\left[L_1^h L_2^k\binom{n}{k}\cdot\sum_{i=0}^{p}\binom{h}{i}\times\cos(h-2i)\theta_1\cdot\sum_{j=0}^{q}\binom{k}{j}\cos(k-2j)\theta_2\right]\right\}. \tag{A5}$$

To obtain CDT $(2f_1-f_2)$, n and k should be odd numbers $(k<n)$. Moreover, $i=p=[(n-k)/2]-1$ and $j=q=(k-1)/2$, so that $(h-2i)\theta_1=2\theta_1$, and $(k-2j)\theta_2=\theta_2$

$$y(2\theta_1,\theta_2)=\sum_{n=0}^{\infty}\left\{\frac{a_n}{2^{n-2}}\sum_{k=1}^{n}\left[L_1^h L_2^k\binom{n}{k}\binom{h}{i}\cos(2\theta_1)\cdot\binom{k}{j}\cos\theta_2\right]\right\}. \tag{A6}$$

Recall that $\cos\alpha\cdot\cos\beta=\frac{1}{2}[\cos(\alpha-\beta)+\cos(\alpha+\beta)]$; thus, the magnitude of the CDT $(2\theta_1-\theta_2)$ is

$$M_{2f_1-f_2}=\left|\sum_{n=3}^{\infty}\left\{\frac{a_n}{2^{n-1}}\sum_{k=1}^{n}\left[L_1^h L_2^k\binom{n}{k}\binom{h}{i}\binom{k}{j}\right]\right\}\right|, \tag{A7}$$

here, the absolute value is adopted for M_{2f1-f2}, since the magnitude cannot be negative; and

$$\binom{n}{k}\binom{h}{i}\binom{k}{j}=n!\left[\left(\frac{h-2}{2}\right)!\left(\frac{h+2}{2}\right)!\left(\frac{k-1}{2}\right)!\left(\frac{k+1}{2}\right)!\right]^{-1}.$$

Replacing a_n with $(1/n!)f^{(n)}(x_0)$, we have

$$M_{2f_1-f_2}=\left|\sum_{n=3}^{\infty}\left\{\frac{f^{(n)}(x_0)}{2^{n-1}n!}\sum_{k=1}^{n}\left[L_1^h L_2^h\binom{n}{k}\binom{h}{i}\binom{k}{j}\right]\right\}\right|, \tag{A8}$$

where $f^{(n)}(x_0)$ is the nth derivative of the original $f_{NL}(x)$ evaluated at OP.

When $L_1<1$, and $L_2<1$, the series in Eq. (A8) always converges. Because $\lim_{n\to\infty}1/(2^{n-1}\cdot n!)=0$, the nth element of the series approaches zero. Thus, the first element of the series is the largest. Let us separate the first element from the rest

$$M_{2f_1-f_2}=L_1^2L_2\left|\frac{3}{4}\frac{f^{(3)}(x_0)}{3!}+\sum_{n=5,7,\ldots}^{\infty}\left[\frac{f^{(n)}(x_0)}{2^{n-1}n!}\times\sum_{k=1,3,\ldots}^{n}L_1^{h-2}L_2^{k-1}\binom{n}{k}\binom{h}{i}\binom{k}{j}\right]\right|. \tag{A9}$$

Thus, if L_1, $L_2\ll 1$, and L_1, $L_2\to 0$, the second part of Eq. (A9)$\to 0$, i.e., the first element is the predominant component in the formation of CDT. So, the M_{2f1-f2} can be approximated as

$$M_{2f_1-f_2}\approx\tfrac{1}{8}L_1^2L_2\cdot|f^{(3)}(x_0)|. \tag{A10}$$

When L_1 and L_2 are fixed, we have

$$M_{2f_1-f_2}\propto|f^{(3)}(x_0)|. \tag{A11}$$

APPENDIX B: DERIVATIVES OF SECOND-ORDER BOLTZMANN FUNCTION

The second-order $f_B(x)$ is generally described as

$$y=\frac{A}{1+e^{bx-c}(1+e^{dx-e})}. \tag{B1}$$

Let $m=1/e^c$, $n=1/e^{c+e}$, and $k=b+d$; the above function can be written as

$$y=\frac{A}{1+me^{bx}+ne^{kx}}=\frac{A}{p}. \tag{B2}$$

1. First derivative

Since $dy/dp=-A/p^2$, and let $bk=dp/dx=bme^{bx}+kne^{kx}$, then the first derivative is

$$y'=\frac{dy}{dx}=\frac{dy}{dp}\frac{dp}{dx}=-A\frac{bk}{p^2}. \tag{B3}$$

2. Second derivative

$$y''=\frac{dy'}{dx}=-A\frac{bk'p^2-bk(p^2)'}{(p^2)^2}. \tag{B4}$$

Let $bk2=bk'=b^2me^{bx}+k^2ne^{kx}$. Since $(p^2)'=2pp'=2p\cdot bk$, the second derivative:

$$y''=-\frac{A}{p^3}(p\cdot bk2-2bk^2). \tag{B5}$$

3. Third derivative

$$y''' = (y'')' = -A\left[\left(\frac{bk2}{p^2}\right)' - 2\left(\frac{bk^2}{p^3}\right)'\right]. \tag{B6}$$

Let $bk3 = (bk2)' = b^3 m e^{bx} + k^3 n e^{kx}$. Since $(bk^2)' = 2bk \cdot bk' = 2bk \cdot bk2$, and $(p^3)' = 3p^2 \cdot p' = 3p^2 \cdot bk$, then

$$\left(\frac{bk2}{p^2}\right)' = \frac{(bk2)' \cdot p^2 - bk2 \cdot (p^2)'}{(p^2)^2}$$

$$= \frac{bk3 \cdot p^2 - bk2 \cdot 2p \cdot bk}{p^4}. \tag{B7}$$

Also, $(bk^2/p^3)' = [(bk^2)' \cdot p^3 - bk^2 \cdot (p^3)']/p^6$; substituted with known parts, it becomes

$$\left(\frac{bk^2}{p^3}\right)' = \frac{bk}{p^4}(2p \cdot bk2 - 3bk^2). \tag{B8}$$

Replacing corresponding parts in Eq. (B6) with Eqs. (B7) and (B8), we have

$$y''' = -A\left[\frac{p^2 \cdot bk3 - 2p \cdot bk \cdot bk2}{p^4} - 2\frac{bk(2p \cdot bk2 - 3bk^2)}{p^4}\right]. \tag{B9}$$

Simplified and rearranged, the third derivative is

$$y''' = -\frac{A}{p^4}(p^2 \cdot bk3 - 6p \cdot bk \cdot bk2 + 6bk^3). \tag{B10}$$

Abdala, C. (**2000**). ''Distortion product otoacoustic emission ($2f1-f2$) amplitude growth in human adults and neonates,'' J. Acoust. Soc. Am. **107**, 446–456.

Avan, P., Magnan, P., Smurzynski, J., Probst, R., and Dancer, A. (**1998**). ''Direct evidence of cubic difference tone propagation by intracochlear acoustic pressure measurements in the guinea pig,'' Eur. J. Neurosci. **10**, 1764–1770.

Bian, L., and Chertoff, M. E. (**2001**). ''Distinguishing cochlear pathophysiology in 4-aminopyridine and furosemide treated ears using a nonlinear systems identification technique,'' J. Acoust. Soc. Am. **109**, 671–685.

Bian, L., and Chertoff, M. E. (**1998**). ''Differentiation of cochlear pathophysiology in ears damaged by salicylate or a pure tone using a nonlinear systems identification technique,'' J. Acoust. Soc. Am. **104**, 2261–2271.

Brown, A. M. (**1987**). ''Acoustic distortion from rodent ears: A comparison of responses from rats, guinea pigs, and gerbils,'' Hear. Res. **31**, 25–38.

Brown, A. M., and Kemp, D. T. (**1984**). ''Suppressibility of the $2f1-f2$ stimulated acoustic emissions in gerbil and man,'' Hear. Res. **13**, 29–38.

Cheatham, M. A., and Dallos, P. (**1994**). ''Stimulus biasing: A comparison between cochlear hair cell and organ of Corti response patterns,'' Hear. Res. **75**, 103–113.

Cheng, J. (**1999**). ''Estimation of active cochlear nonlinearity by multicomponent distortion-product otoacoustic emissions,'' Acustica **85**, 721–727.

Chertoff, M. E., Steele, T. C., and Bian, L. (**1997**). ''Characterizing cochlear mechano-electric transduction in ears damaged with pure tones,'' J. Acoust. Soc. Am. **102**, 441–450.

Chertoff, M. E., Steele, T., Ator, G. A., and Bian, L. (**1996**). ''Characterizing cochlear mechano-electric transduction using a nonlinear system identification procedure,'' J. Acoust. Soc. Am. **100**, 3741–3753.

Corey, D. P., and Hudspeth, A. J. (**1983**). ''Kinetics of the receptor current in bullfrog saccular hair cells,'' J. Neurosci. **3**, 962–976.

Crawford, A. C., Evans, M. G., and Fettiplace, R. (**1989**). ''Activation and adaptation of transducer currents in turtle hair cells,'' J. Physiol. (London) **419**, 405–434.

Dallos, P. (**1986**). ''Neurobiology of cochlear inner and outer hair cells: Intracellular recordings,'' Hear. Res. **22**, 185–198.

Dallos, P., Evans, B. N., and Hallworth, R. (**1991**). ''Nature of the motor element in electrokinetic shape changes of cochlear outer hair cells,'' Nature (London) **350**, 155–157.

Engebretson, A. M., and Eldredge, D. H. (**1968**). ''Model for the nonlinear characteristics of cochlear potentials,'' J. Acoust. Soc. Am. **44**, 548–554.

Evans, B. N., Hallworth, R., and Dallos, P. (**1991**). ''Outer hair cell electromotility: The sensitivity and vulnerability of the dc component,'' Hear. Res. **52**, 288–304.

Fahey, P. F., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**2000**). ''Nonlinear interaction that could explain distortion product interference response areas,'' J. Acoust. Soc. Am. **108**, 1786–1802.

Frank, G., and Kössl, M. (**1996**). ''The acoustic two-tone distortions $f2-f2$ and $f2-f1$ and their possible relation to changes in the operating point of the cochlear amplifier,'' Hear. Res. **98**, 104–115.

Frank, G., and Kössl, M. (**1997**). ''Acoustic and electrical biasing of the cochlear partition. Effects on the acoustic two tone distortions f_2-f_1 and $2f_1-f_2$,'' Hear. Res. **113**, 57–68.

Geisler, C. D., and Nuttall, A. L. (**1997**). ''Two-tone suppression of basilar membrane vibrations in the base of the guinea pig cochlea using 'low-side' suppressors,'' J. Acoust. Soc. Am. **102**, 430–440.

He, N., and Schmiedt, R. A. (**1997**). ''Fine structure of the $2f1-f2$ acoustic distortion products: Effects of primary level and frequency ratios,'' J. Acoust. Soc. Am. **101**, 3554–3565.

Holton, T., and Hudspeth, A. J. (**1986**). ''The transduction channel of hair cells from the bull-frog characterized by noise analysis,'' J. Physiol. (London) **375**, 195–227.

Kirk, D. L., and Patuzzi, R. B. (**1997**). ''Transient changes in cochlear potentials and DPOAEs after low-frequency tones: The 'two-minute bounce' revisited,'' Hear. Res. **112**, 49–68.

Kirk, D. L., and Yates, G. K. (**1998**). ''Enhancement of electrically evoked oto-acoustic emissions associated with low-frequency stimulus bias of the basilar membrane towards scala vestibuli,'' J. Acoust. Soc. Am. **104**, 1544–1554.

Knight, R. D., and Kemp, D. T. (**2000**). ''Indications of different distortion product otoacoustic emission mechanisms from a detailed f_1, f_2 area study,'' J. Acoust. Soc. Am. **107**, 457–473.

Konishi, T., Butler, R. A., and Fernandez, C. (**1961**). ''Effects of anoxia on cochlear potentials,'' J. Acoust. Soc. Am. **33**, 349–356.

Kros, C. J., Rüsch, A. R., and Richardson, G. P. (**1992**). ''Mechano-electrical transducer currents in hair cells of the cultured neonatal mouse cochlea,'' Proc. R. Soc. London, Ser. B **421**, 185–193.

Lonsbury-Martin, B. L., Martin, G. K., Probst, R., and Coats, A. C. (**1987**). ''Acoustic distortion products in rabbit ear canal. I. Basic features and physiological vulnerability,'' Hear. Res. **28**, 173–190.

Lonsbury-Martin, B. L., Harris, F. P., Stagner, B. B., Hawkins, M. D., and Martin, G. K. (**1990**). ''Distortion product emissions in humans. I. Basic properties in normally hearing subjects,'' Ann. Otol. Rhinol. Laryngol. Suppl. **147**, 3–14.

Lukashkin, A. N., and Russell, I. J. (**1998**). ''A descriptive model of the receptor potential nonlinearities generated by the hair cell mechanoelectrical transducer,'' J. Acoust. Soc. Am. **103**, 973–980.

Lukashkin, A. N., and Russell, I. J. (**1999**). ''Analysis of the f_2-f_1 and $2f_1-f_2$ distortion components generated by the hair cell mechanoelectrical transducer: Dependence on the amplitudes of the primaries and feedback gain,'' J. Acoust. Soc. Am. **106**, 2661–2668.

Martin, G. K., Lonsbury-Martin, B. L., Probst, R., Scheinin, S. A., and Coats, A. C. (**1987**). ''Acoustic distortion products in rabbit ear canal. II. Sites of origin revealed by suppression contours and pure-tone exposures,'' Hear. Res. **28**, 191–208.

Mills, D. M. (**1997**). ''Interpretation of distortion product otoacoustic emission measurements. I. Two stimulus tones,'' J. Acoust. Soc. Am. **102**, 413–429.

Mills, D. M., and Rubel, E. W. (**1994**). ''Variation of distortion product otoacoustic emissions with furosemide injection,'' Hear. Res. **77**, 183–199.

Mills, D. M., Norton, S. J., and Rubel, E. W. (**1993**). ''Vulnerability and adaptation of distortion product otoacoustic emissions to endocochlear potential variation,'' J. Acoust. Soc. Am. **94**, 2108–2122.

Mom, T., Bonfils, P., Gilain, L., and Avan, P. (**2001**). ''Origin of cubic difference tones generated by high-intensity stimuli: Effect of ischemia and auditory fatigue on the gerbil cochlea,'' J. Acoust. Soc. Am. **110**, 1477–1488.

Nelson, D. A., and Kimberley, B. P. (**1992**). ''Distortion-product emissions

and auditory sensitivity in human ears with normal hearing and cochlear hearing loss,'' J. Speech Hear. Res. **35**, 1142–1159.
Nieder, P., and Nieder, I. (**1971**). ''Determination of microphonic generator transfer characteristic from modulation data,'' J. Acoust. Soc. Am. **49**, 478–492.
Nilsen, K. E., and Russell, I. J. (**2000**). ''The spatial and temporal representation of a tone on the guinea pig basilar membrane,'' Proc. Natl. Acad. Sci. U.S.A. **97**, 11751–11758.
Patuzzi, R., and Moleirinho, A. (**1998**). ''Automatic monitoring of mechano-electrical transduction in the guinea pig cochlea,'' Hear. Res. **125**, 1–16.
Popelka, G. R., Osterhammel, P. A., Nielsen, L. H., and Rasmussen, A. N. (**1993**). ''Growth of distortion product otoacoustic emissions with primary-tone level in humans,'' Hear. Res. **71**, 12–22.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Russell, I. J., and Kössl, M. (**1992a**). ''Modulation of hair cell voltage responses to tones by low-frequency biasing of the basilar membrane in the guinea pig cochlea,'' J. Neurosci. **12**, 1587–1601.
Russell, I. J., and Kössl, M. (**1992b**). ''Voltage responses to tones of outer hair cells in the basal turn of the guinea-pig cochlea: significance for electromotility and desensitization,'' Proc. R. Soc. London, Ser. B **247**, 97–105.
Russell, I. J., Cody, A. R., and Richardson, G. P. (**1986**). ''The responses of inner and outer hair cells in the basal turn of the guinea-pig cochlea and in the mouse cochlea grown *in vitro*,'' Hear. Res. **22**, 199–216.
Santos-Sacchi, J. (**1989**). ''Asymmetry in voltage-dependent movements of isolated outer hair cells from the organ of Corti,'' J. Neurosci. **9**, 2954–2962.
Santos-Sacchi, J. (**1991**). ''Reversible inhibition of voltage-dependent outer hair cell motility and capacitance,'' J. Neurosci. **11**, 3096–3110.
Santos-Sacchi, J. (**1993**). ''Harmonics of outer hair cell motility,'' Biophys. J. **65**, 2217–2227.
Scholz, G., Hirschfelder, A., Marquardt, T., Hensel, J., and Mrowinski, D. (**1999**). ''Low-frequency modulation of the $2f_1-f_2$ distortion product otoacoustic emissions in the human ears,'' Hear. Res. **130**, 189–196.
Takahashi, S., and Santos-Sacchi, J. (**1999**). ''Distortion component analysis of outer hair cell motility-related gating charge,'' J. Membr. Biol. **169**, 199–207.
van Emst, M. G., Klis, S. F., and Smoorenburg, G. F. (**1997**). ''Identification of the nonlinearity governing even-order distortion products in the cochlear potentials,'' Hear. Res. **114**, 93–101.
Weiss, T. F., and Leong, R. (**1985**). ''A model for signal transmission in an ear having hair cells with free-standing stereocilia. IV. Mechanoelectric transduction stage,'' Hear. Res. **20**, 175–195.
Yates, G. K., and Kirk, D. L. (**1998**). ''Cochlear electrically evoked emissions modulated by mechanical transduction channels,'' J. Neurosci. **15**, 1996–2003.

Stimulus set effects in the similarity ratings of unfamiliar complex sounds

Prudence Allen and Susan Scollie
National Centre for Audiology, School of Communication Sciences and Disorders, University of Western Ontario, Elborn College, 1201 Western Road, London, Ontario N6G 1H1, Canada

(Received 16 November 2000; revised 2 April 2002; accepted 5 April 2002)

This study examined the effect of the distribution of values for various acoustic properties on the similarity ratings obtained in a paired-comparison study of complex sounds. Listeners rated the similarity of tones and tone complexes in one of four sets. Across the four sets both component frequency and the number of components in the stimuli were drawn from distributions that were either categorical or noncategorical. Other acoustic properties were distributed similarly in all sets. Solutions suggested that frequency was most important to the ratings, followed by the number of components in each critical band. Results suggest that when the values of an acoustic property are distributed categorically, its importance to similarity ratings increases relative to that observed with a less categorical distribution. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1482075]

PACS numbers: 43.66.Ba, 43.66.Fe, 43.66.Lj [MRL]

I. INTRODUCTION

Natural sounds are generally complex, differing in multiple acoustic dimensions, some of which may be important for discrimination and identification and others that may not. To learn new sounds, differentiating one from another and assigning meaning to each, listeners must weight acoustic information such that more important dimensions are weighted more heavily than those that are less important. Insight into how listeners weight acoustic information, that is, whether weightings are relatively fixed or may vary with factors such as the set of stimuli in question, the age of the listener, or the listeners prior experience, are important to the understanding of auditory learning and initial stages of encoding. In this study, one of these factors, the structure of the stimulus set, is examined.

The manner in which listeners weight acoustic information can be studied using paired comparisons procedures and Multidimensional Scaling Analyses (MDS). Listeners are asked to rate the similarity of pairs of stimuli within a set. The MDS analysis uses the similarity ratings to derive a stimulus space of relatively few dimensions (generally 2 to 3) in which stimuli that were perceived as similar are placed close together and those perceived as dissimilar are placed far apart. The dimensions of the derived space are theorized to correspond to stimulus parameters used by the listeners in evaluating the stimuli. The relative importance of the dimensions (and the stimulus parameters corresponding to them) is indicated by the dimension weights. When individual differences models are used for the MDS analysis (e.g., INDSCAL, Carroll and Chang, 1970), it is possible to determine the extent to which individual listeners' weightings conform to those derived from the group data.

MDS techniques have been used to evaluate observers' perceptions of a variety of complex stimuli but only a few studies have used auditory stimuli. Allen and Bond (1997) asked 7- and 10-year-old children and adults to rate the similarity of a set of unfamiliar, complex sounds that was constructed by varying a very large number of acoustic parameters. The purpose of the study was to determine which acoustic parameters would be more important to listeners and whether there would be age-related differences in relative importance. The sounds varied in multiple parameters including mean frequency, overall bandwidth, pitch, the number of peaks in the amplitude spectrum, the number of frequency components, and the shape of the temporal envelope. Although many acoustic parameters had been manipulated in stimulus construction, only a small number were required to account for the majority of the variance in the ratings. These included relative periodicity, the number of peaks in the amplitude spectra, and mean frequency. The authors suggested that the results may have been influenced by the manner in which the acoustic parameter values were distributed in the stimulus set. That is, the two acoustic properties that appeared most important to the ratings, relative periodicity and number of spectral peaks, were those that could take on only one of two to three values (i.e., they showed a categorical distribution of values). The organization of stimuli within the derived space was consistent with grouping based upon these two parameters. The third most significant acoustic parameter, frequency, was distributed relatively continuously across the stimulus set and in the derived solution it was seen that the clusters of stimuli formed on the basis of periodicity and spectral peaks were each organized according to mean frequency.

Some theories of similarity have suggested a link between feature salience and the extent to which the feature is useful for the classification of stimuli within the set (e.g., Tversky, 1977). The argument is that, when presented with a large stimulus set, a listener/observer will attempt to reduce the information processing demands of the large number of items by organizing the stimuli into smaller subsets of items. Features that are useful for this organization will appear to be most salient (i.e., correlate most significantly with the dimensions of a derived space in an MDS analysis). As structure is added to the stimulus set by the listener, the similarity ratings (and the measures of salience derived from them) may relate

less to the physical characteristics of the stimuli themselves and more to the structure that the listener has placed on the set (Nosofsky, 1986; Sjoberg and Thorslund, 1979; Torgerson, 1965; Tversky, 1977). This raises the question whether, in the study by Allen and Bond (1997), the parameters that were most important to the similarity ratings were those that were more important in general or simply appeared most important in the rating of that set of stimuli because their values were distributed categorically and therefore useful for the organization of stimuli within the stimulus set. Frequency may have been less significant than periodicity and spectral peaks simply because of its distribution of values within the set.

Examination of other MDS studies finds some support for this. For example, Christensen and Humes (1996) asked listeners to evaluate the similarity of a highly structured set of eight brief tonal complexes that were formed from a combination of acoustic parameters that each took on one of only two possible values. These parameters included harmonicity, spectral shape, and temporal envelope. The solution derived from the similarity ratings suggested that these parameters were highly important to the ratings and within the derived space the stimuli were placed in clusters organized according to these parameters. Similar results were observed in a later study using 27 synthesized speechlike sounds (Christensen and Humes, 1997) that varied in fricative spectrum center frequency, the slope of the frequency transition, and the duration of a temporal gap. Thus, categorically distributed acoustic features often appear to be important in MDS solutions. However a systematic examination of feature distribution patterns on relative salience has not been reported.

The purpose of this study was to more thoroughly examine the effect of the distribution of parameter values within stimulus sets on similarity ratings and subsequent estimates of salience derived from them. The questions addressed were (a) what acoustic parameter(s) would be most salient to listeners, (b) would salience estimates be relatively consistent across stimulus sets, and (c) would relative salience vary with the manner in which the values of a stimulus parameter are distributed.

II. METHOD

A. Listeners

Twenty-four young adult listeners (mean age $=27$ years, range$=23$ to 40 years) were recruited from the university community. All were required to pass an audiometric screening with thresholds at or below 20 dB HL (ANSI, 1989) from 500 to 4000 Hz, and normal tympanograms. All read a letter of information and provided signed consent before beginning the study. Five listeners were male and 19 were female.

B. Stimuli

Stimuli were pure tones or tonal complexes consisting of equal amplitude components drawn from 3 high or 3 low frequency bands and containing up to nine components. Both the frequency of the components (low versus high) and the number of components (few versus many) were distributed in a manner that made them relatively continuous or categorical. Overall bandwidth and the number of frequency bands in the stimuli were distributed similarly across the sets. Four sets were created each of which was used in one condition of the study.

TABLE I. Frequency bands from which components were drawn. Low (L1 to L3) and high (H1 to H3) frequency bands are each approximately one critical band in width and separated from adjacent bands by one critical band. The three low bands are separated from the three high bands by approximately one critical band in conditions 1 and 2 and three critical bands in conditions 3 and 4.

	Conditions 1 and 2	Conditions 3 and 4
L1	553–669	456–555
L2	819–961	651–777
L3	1204–1391	937–1080
H1	1795–2020	1880–1912
H2	3147–3519	2654–3048
H3	3725–4346	3722–4350

(1) Number and frequency of components distributed continuously.
(2) Number of components distributed categorically and component frequency distributed continuously.
(3) Number of components distributed continuously and component frequency distributed categorically.
(4) Both number and frequency of components distributed categorically.

1. Selection of component frequencies

The frequency bands from which components were drawn are shown in Table I. Half of the stimuli in each condition consisted of components drawn from one to three of the high frequency bands (H1, H2, and H3) and half from the low (L1, L2, and L3). In all conditions the frequency bands were approximately one critical band in width (Scharf, 1970) and equally spaced on a log scale with one critical band between them. In conditions 1 and 2 the separation between the three low and three high frequency bands was also one critical band making the distribution of frequency bands from which the components were drawn continuous. In conditions 3 and 4 the three low frequency bands were separated from the three high frequency bands by approximately three critical bands, creating a larger, more categorical distinction, between high and low frequency stimuli.

2. Number of components and number of components per frequency band

The number of components in each stimulus and the frequency bands from which those components were drawn are shown in Table II. In all four conditions, approximately half of the stimuli (equally distributed among the stimuli from both the high and low frequency bands) were composed of components drawn from only one frequency band. The other half were drawn from two or three bands. This produced a roughly equal number of stimuli with single and multiple peaks in their amplitude spectra.

Stimuli contained a maximum of nine components that were evenly distributed across the number of frequency

TABLE II. Stimulus composition for the 18 stimuli in conditions 1–4. The number in parentheses indicates the number of components in the stimulus. L1, L2, L3, H1, H2, and H3 indicate the frequency bands from which these components were drawn. When a stimulus contained components drawn from more than one frequency band, they were distributed equally across the bands. See Table I for the ranges of frequencies included in each band.

	Condition 1	Condition 2	Condition 3	Condition 4
1	(1) L1	(1) L1	(1) L1	(1) L1
2	(4) L1	(5) L1	(4) L1	(5) L1
3	(2) L2	(1) L2	(2) L2	(2) L2
4	(3) L3	(1) L3	(3) L3	(1) L3
5	(8) L3	(7) L3	(8) L3	(7) L3
6	(4) L1, L3	(2) L1, L3	(4) L1, L3	(2) L1, L3
7	(3) L1, L2, L3	(6) L1, L3	(3) L1, L2, L3	(6) L1, L3
8	(6) L1, L2, L3	(6) L1, L2, L3	(6) L1, L2, L3	(6) L1, L2, L3
9	(9) L1, L2, L3	(9) L1, L2, L3	(9) L1, L2, L3	(9) L1, L2, L3
10	(1) H1	(1) H1	(1) H1	(1) H1
11	(5) H1	(6) H1	(5) H1	(6) H1
12	(2) H2	(2) H2	(2) H2	(2) H2
13	(3) H3	(1) H3	(3) H3	(1) H3
14	(9) H3	(8) H3	(9) H3	(8) H3
15	(2) H1, H3	(2) H1, H3	(2) H1, H3	(2) H1, H3
16	(2) H1, H3	(6) H1, H3	(4) H1, H3	(6) H1, H3
17	(6) H1, H3	(6) H1, H2, H3	(6) H1, H3	(6) H1, H2, H3
18	(6) H1, H2, H3	(9) H1, H2, H3	(6) H1, H2, H3	(9) H1, H2, H3

bands within the stimulus. In this way the temporal fine structure would be constant in all spectral regions. Thus, in conditions 1 and 3 for which the number of components was noncategorical, a stimulus with components drawn from one frequency band could contain up to nine components; a stimulus with two frequency bands would contain 2, 4, or 6 components, and a stimulus with components drawn from three frequency bands could contain 3, 6, or 9 components. In conditions 2 and 4 the number of components was restricted such that no stimuli could contain three or four components. The middle third of the range from 1 to 9 components, in log units, was eliminated to produce more categorical distinction between stimuli with few versus many components.[1] Thus, a signal with components drawn from a single frequency band could have one, two, or five to nine components. A stimulus with components drawn from two spectral bands could have two, six, or eight components, and a stimulus with components drawn from three spectral bands could have six or nine components.

The exact frequency of the components drawn from each frequency band was selected from one of nine frequencies spaced at equal log intervals within each band. When a band was selected for a stimulus the required number of components were selected randomly without replacement from the nine available. Thus, a single frequency component could be drawn only once in the construction of an individual stimulus but could be drawn for inclusion in multiple stimuli.

As is summarized in Table II, each condition has equal numbers of low and high frequency stimuli, and approximately equal numbers of stimuli with components in single and multiple (2 to 3) frequency bands. The overall range of frequencies is constant across conditions (approximately 500 to 4000 Hz) and each condition includes approximately four to six stimuli with one to three components (such that the individual components would likely be resolved by the listener), or greater than three components (such that individual components are less likely to be resolved) in each frequency band. The only differences in the conditions is the separation between the high and low frequency distributions which is greater in conditions 3 and 4 than 1 and 2; and in the number of components in each stimulus with conditions 2 and 4 having no stimuli with only three or four components.

In all conditions the stimuli were 430 ms in duration including 10 ms cosine-squared on and off ramps. Stimuli were digitally generated at a 25 kHz sampling rate, converted to analog form by an Ariel DSP-16 D/A converter, low pass filtered at 8 kHz, attenuated by custom-made attenuators, and presented over TDH-49 headphones at an overall level that varied randomly over a 10 dB range centered at 65 dB SPL.

C. Procedure

Six of the 24 listeners were assigned to each of the four conditions, in random order. In this way, potential bias in similarity ratings resulting from prior exposure to another stimulus set could be avoided. Each condition consisted of 171 trials (all possible pairs were evaluated only once and identical pairs were included), completed in blocks of approximately 25. Listeners were able to take breaks between blocks if needed. All testing was completed in a sound treated room (IAC) at the university.

The procedure was developed to be suitable for use with both adults and children. On each trial, listeners were shown two identical images on the computer screen, presented in sequence and separated by a 500 ms interval. Each image was approximately 1.5″ wide and 3″ high. The first image was fixed at the left of the screen and the second image was randomly placed between the left image and the right side of the screen. Coincident with the presentation of each image the two sounds of the pair were heard, one in each ear. To indicate the similarity of the sounds the listeners could move the right most picture either closer to or farther from the left

TABLE III. Total variance accounted for by the fitted MDS solution and stress values associated with those fits for each condition.

Condition	Variance accounted for	Stress
1: Both noncategorical	62.70	0.22
2: Components categorical	51.12	0.24
3: Frequency categorical	64.05	0.22
4: Both categorical	58.93	0.22

most picture, which was fixed. The distance between the images, in percent of the total available, was taken as the dissimilarity measure, thus minimum and maximum dissimilarity ratings were 0 and 100. Listeners were instructed to use the computer mouse to move the images close together if they were similar and further apart if they were dissimilar. Once they had moved the image, the sounds were repeated and the listener was allowed to adjust their response if desired. Listeners indicated their satisfaction with their response and their readiness for the next trial by pressing the enter key. The screen was then cleared and a different pair of images and sounds was presented. (The option to readjust responses was not used by the adults although it was available.)

Training on the task consisted of a brief practice condition of six trials with a set of three stimuli: a four-tone complex, a narrow-band noise centered at 500 Hz, and a wide band noise extending from 0 to 8000 Hz with a spectral notch 40 dB deep, centered at 4000 Hz. Following the practice, the listeners were familiarized with the range of stimuli they would be asked to judge through a single presentation of the experimental stimuli. They were told that these were the stimuli that they would be asked to rate, in pairs. Data collection was completed for each listener in one to two sessions per listener totaling less than 1 hour.

D. MDS analysis

Similarity measures obtained for each condition were submitted to a multidimensional scaling analysis that used an alternating least squares (ALSCAL) fitting algorithm and the individual differences (INDSCAL) model of analysis (Carroll and Chang, 1970). The proximity measures were assumed to be ordinal and tied scores were left tied. A three dimensional solution is reported. Variance accounted for by the fitted solution (R^2) and stress values for each of the four conditions are shown in Table III.

To assist in defining the acoustic parameters that may have been represented in the derived solutions, the coordinates of each stimulus in the MDS derived space were submitted as dependent variables in a simultaneous multiple regression analysis (Fox, 1983). Independent variables were measures of the acoustic characteristics of the stimuli. In the initial analysis the independent variables included average frequency (in log units), the overall number of components, the number of frequency bands from which components were drawn, the number of components falling in each critical band, bandwidth (in log units), and a measure of overall periodicity (the inverse of the variance in the intervals between successive zero crossings in the time domain wave form). Bandwidth and overall periodicity did not contribute significantly to the regression equation and were excluded from the final analysis. The overall number of components was significantly correlated with the number of components in each critical band (Pearson correlation $=0.787$ for all conditions) and with the number of bands (correlation$=0.465$ for all conditions), likely because of the manner in which the stimuli were constructed. To eliminate suppression effects, correlated parameters could not be included in the analysis. It was decided that the final regression analysis should include parameters that were interpretable and consistent with the presumed function of the auditory periphery. Because the auditory system performs a spectral analysis at the initial stages of encoding with other characteristics, such as temporal structure, evaluated within critical bands, it was decided to retain the number of components per critical band rather than the number of components overall although either parameter produced similar results in accounting for the derived dimensions. The psychological percept associated with the number of components in each critical band is likely most closely associated with roughness. Thus, the final analyses included only average frequency, number of frequency bands, and the number of components in each frequency band.

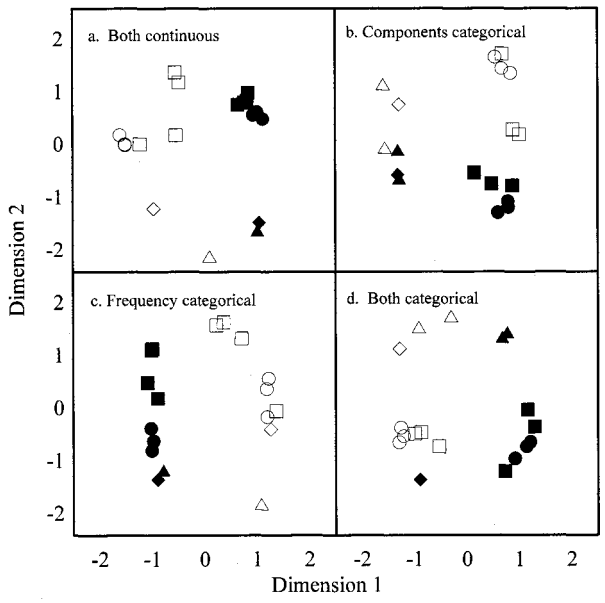


FIG. 1. MDS solutions derived from the dissimilarity data in each of the four conditions are shown in panels (a)–(d). Dimension 2 weights are shown on the ordinate and dimension 1 weights on the abscissa. Stimuli drawn from the low and high frequency distributions are shown with open and filled symbols, respectively. Single band stimuli with only one component (i.e., pure tones) are shown by triangles. Stimuli with components drawn from more than one frequency band but that contain only one component in each band are shown by the diamonds. Stimuli containing more than one component in each frequency band but with components in only one band are shown by the squares. Circles indicate stimuli containing more than one component in each frequency band and have components drawn from more than one band.

III. RESULTS

MDS solutions derived from the similarity data in each of the four conditions are shown in the four panels of Figs. 1 and 2. Figure 1 shows dimension 2 as a function of dimen-

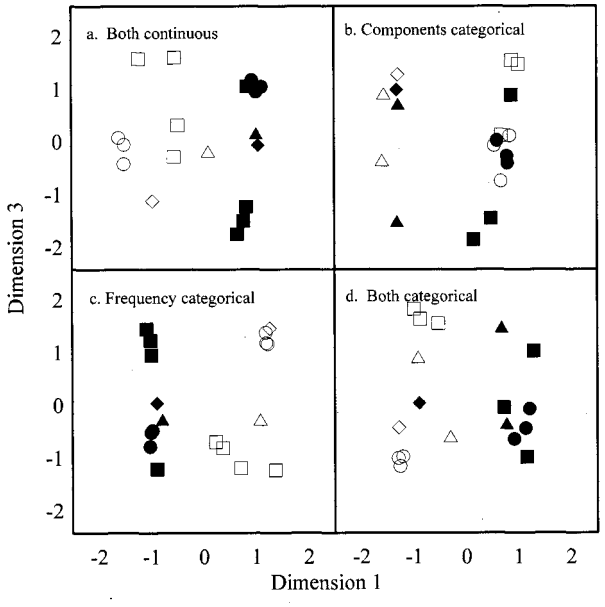


FIG. 2. MDS solutions derived from the dissimilarity data in each of the four conditions are shown in panels (a)–(d). Dimension 1 weights are shown on the ordinate and dimension 3 weights on the abscissa. Stimuli drawn from the low and high frequency distributions are shown with open and filled symbols, respectively. Single band stimuli with only one component (i.e., pure tones) are shown by triangles. Stimuli with components drawn from more than one frequency bands but that contain only one component in each band are shown by the diamonds. Stimuli containing more than one component in each frequency band but with components in only one band are shown by the squares. Circles indicate stimuli containing more than one component in each frequency band and have components drawn from more than one band.

sion 1 and Fig. 2 shows dimension 3 plotted against dimension 1. Open and filled symbols show stimuli drawn from the lower and higher frequency bands, respectively. Triangles indicate stimuli with one component, diamonds indicate stimuli with components drawn from more than one frequency band but containing only one component in each band, squares indicate stimuli with multiple components contained entirely within one frequency band, and circles indicate stimuli with multiple components in each of several frequency bands.

Condition 1: Component frequency and number distributed noncategorically. Figures 1(a) and 2(a) show the MDS configuration for stimuli with both component frequency and number distributed continuously. The MDS model accounted for 62.7% of the variance in the similarity ratings. Dimension 1 accounted for 31.3% of the variance. The multiple regression analysis suggested that the three acoustic parameters were good predictors of dimension 1 placement of stimuli ($R^2=0.826$). The standardized coefficient was highest for mean frequency ($\beta=0.855$), reflected in the separation of open and filled symbols along dimension 1 in Fig. 1(a). There was also a small, but statistically significant, contribution made by the number of components in each frequency band ($\beta=-0.341$) and number of bands ($\beta=-0.313$). Dimension 2 accounted for 20.7% of the variance in the ratings. The multiple regression analysis produced an $R^2=0.701$ with the highest coefficient associated with the number of components/band, $\beta=0.804$. This can be seen in the separation of circles and squares, representing stimuli with more than one component in each frequency band from the diamonds and triangles, representing stimuli with only one component in each band. Dimension 3, shown in Fig. 2(a), accounted for only 10.7% of the variance in the solution and values were not well predicted by the combination of acoustic parameters ($R^2=0.086$).

Condition 2: Frequency distributed noncategorically and number of components distributed categorically. Figures 1(b) and 2(b) show the fitted three-dimensional space when frequency was distributed continuously but the number of components was distributed categorically. The solution accounted for 51.1% of the variance in the data, slightly lower than was observed in condition 1. Dimension 1 accounted for 21.9% of this variance. A separation can be seen between the stimuli with one component/band (diamonds and triangles) and those with more components/band (circles and squares). The regression analysis with the three predictors produced an R^2 of 0.724, with the most significant contribution made by components/band. There was also a small contribution made by the number of bands ($\beta=0.796$ and 0.332, for components/band and number of bands, respectively). The contribution made by the number of bands may reflect the importance of spectral shape information (each band from which components were drawn would represent one peak in the amplitude spectrum of the sound), along with the number of components/band, or may be an artifact of the stimulus set structure, i.e., none of the stimuli with few components/band had components drawn from more than two frequency bands. A fairly categorical distribution of stimuli can be seen along dimension 1, consistent with the categorical distribution of the number of components and therefore the number of components/band. Dimension 2 accounted for 20.2% of the variance in the data and stimulus placement was best predicted by stimulus frequency (R^2 for the regression$=0.73$, $\beta=-0.854$ for frequency). Unlike the placement of stimuli along dimension 1, placement along dimension 2 was relatively continuous. Dimension 3, shown in Fig. 2(b) accounted for very little of the variance in the data (9%) and stimulus placement was poorly predicted by the set of acoustic parameters, consistent with the results for condition 1 ($R^2=0.229$).

Condition 3: Frequency distributed categorically and number of components distributed noncategorically. Figures 1(c) and 2(c) show the MDS solution for stimuli with component frequencies distributed categorically and number of components distributed continuously. Overall the solution accounted for 64% of the variance in the similarity ratings, similar to condition 1. Dimension 1 accounted for 39.4% of the variance. The regression analysis suggested that placement of stimuli along dimension 1 was best predicted by stimulus frequency with a weaker, but significant, contribution made by the number of frequency bands (total regression $R^2=0.960$, $\beta=-0.963$ and 0.150 for frequency and number of bands, respectively). Dimension 2 correlated with the number of components/band (total $R^2=0.484$, $\beta=0.676$ for components/band) and accounted for 13% of the variance in the overall solution, the smallest proportion of variance attributable to components/band in the four conditions). The distribution of stimuli along dimension 2 was

much more continuous than dimension 1 and the stimuli with fewer components/band (triangles and diamonds) were separated only slightly from the stimuli with more components/band (squares and circles). Consistent with the previous two conditions, dimension 3 [Fig. 2(c)] accounted for 11.6% of the variance in the data and showed no significant correlation with any of the acoustic parameters ($R^2=0.186$). The solution suggested that when frequency was distributed categorically, its contribution to the ratings was greater (i.e., its relative weight in the overall solution increased) than when distributed continuously.

Condition 4: Frequency and number of components distributed categorically. Figures 1(d) and 2(d) show the stimulus configuration for condition 4, in which both frequency and the number of components were distributed categorically. The MDS model accounts for 58.7% of the variance in the similarity measures. Dimension 1 accounted for 31.9% of the variance. The stimulus placement was best predicted by stimulus frequency (total $R^2=0.912$, $\beta=0.948$ for frequency). Dimension 2 accounted for 17.2% of the variance in the ratings and placement of stimuli along dimension 2 was best predicted by the number of components/band. There was also a small contribution made by the number of bands (total $R^2=0.830$, $\beta=-0.851$ and -0.359 for components/band and number of bands, respectively). Dimension 3 [Fig. 2(d)] accounted for only 9.7% of the variance in the ratings but was significantly correlated with the number of frequency bands in the stimuli (total $R^2=0.422$, $\beta=-0.592$ for bands).

Individual differences. Group solutions provided information regarding how feature salience varied with the construction of the stimulus set, but it is important to also evaluate individual differences. This is provided by the INDSCAL analysis in the form of individual subject weights. In general, most listeners weighted the dimensions consistently with the group solution. In conditions 1 and 4 for which the distribution patterns of the most salient acoustic features, corresponding to spectral and temporal information in the signals, were similar, half of the listeners weighted dimension 1 most heavily and half weighted dimensions 1 and 2 roughly the same. In condition 3, for which frequency was distributed categorically and the number of components was distributed noncategorically, the group solution showed that dimension 1, coding frequency, accounted for approximately 10% more variance than did dimension 2, coding periodicity in the signals. In this condition all six of the listeners placed greater weight on dimension 1 than dimension 2 producing the largest degree of homogeneity in the individual listeners responses. In contrast, in condition 2 for which the number of components was categorical, the group solution suggested listeners weighted dimensions 1 and 2 information equally and this was reflected in the individual weights of five of the six listeners. Thus, the distribution of feature values not only affected the weighting of those features in the group solution but also the homogeneity of listeners responses.

IV. DISCUSSION

The goals of this study were to evaluate which acoustic characteristics would be most important to listeners when rating the similarity of stimuli in very similar sets of unfamiliar, brief duration, sounds, and to evaluate whether this importance would change with subtle changes in the distribution of values of an acoustic parameter. The distribution of two parameters, mean frequency and the number of components were manipulated. Because of the rules for the construction of stimuli, the number of components was constrained by the number of frequency bands from which components could be drawn (corresponding to the number of peaks in the amplitude spectra). To keep the temporal fine structure constant across frequency regions it was required that the number of components be equal in each frequency band. Thus, the number of components and the number of components/critical band were highly correlated. While either could be discussed, it was decided to evaluate the parameter that coded the number of components/critical band rather than the number of components overall as that parameter appeared to be more interpretable from the perspective of auditory system functioning than the number of components overall. Thus, discussion of components/critical band and components overall would be similar in terms of their contributions to the analyses, but one is more meaningful than the other for understanding auditory system function.

The absolute values of the standardized regression coefficients for each predictor in the multiple regression analysis are shown for each condition by the filled symbols in each panel of Fig. 3. Also shown is the total variance accounted for in the solution by each of the derived dimensions (open symbols). In general, results suggested that frequency accounted for the largest proportion of the variance in the ratings, best predicting placement of stimuli along dimension 1 of the derived space in three of the four conditions and correlating most highly with dimension 2 of the remaining condition (condition 2). The number of components in each critical band accounted for the second largest proportion of variance, correlating most highly with dimension 2 weights in conditions 1, 3, and 4 and dimension 1 in condition 2. The number of frequency bands from which components were selected also contributed to the ratings although to a much lesser extent. The similarity across conditions is not surprising given that each condition contained tones and tonal complexes drawn from distributions of acoustic parameter values that varied over similar ranges.

Changes in the relative importance of the various parameters with changes in distribution patterns can be seen by examining differences in the extent to which the parameters contributed to the similarity ratings across the four conditions. An acoustic feature that correlates highly with a dimension of the solution that accounts for a large amount of variance in the ratings is considered important. Using this definition, the two most important acoustic parameters in all four conditions were frequency and number of components/critical band, features whose values were distributed categorically. Frequency appeared to be the most important acoustic property across sets in general and, when distributed categorically, its importance was at its highest, especially when the other most important parameter, the number of components/critical band, was distributed continuously (condition 3). It was only when frequency was distributed con-

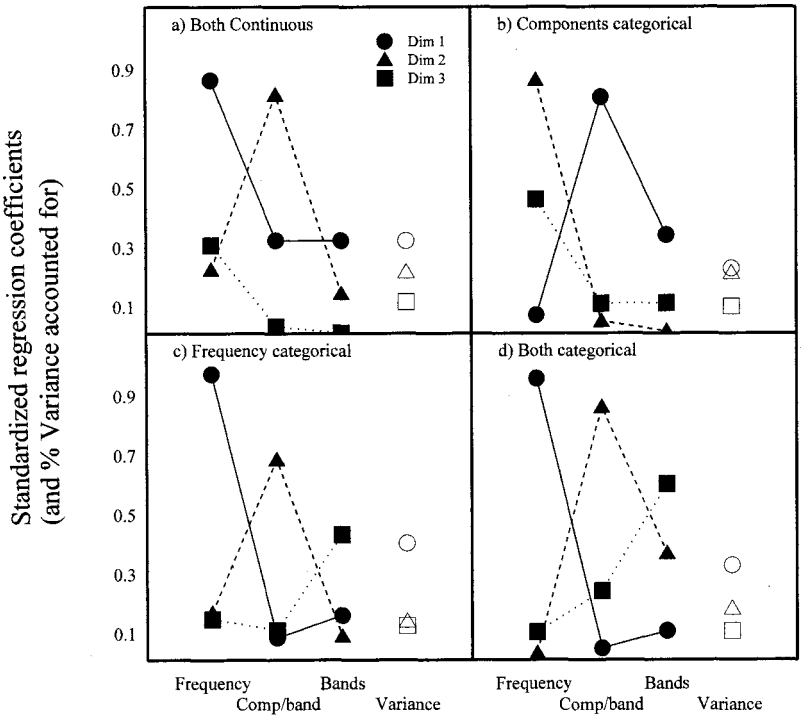


FIG. 3. Results of the multiple regression analysis in which dimension weights were dependent variables and stimulus properties were independent variables. Filled circles, triangles, and squares are used to indicate the absolute values of the obtained standardized regression coefficients for dimensions 1–3, respectively. The open circles, triangles, and squares show the total variance accounted for in the solution by dimensions 1–3, respectively.

tinuously, possibly weakening its salience, and the number of components was distributed categorically thereby increasing its salience, that the balance between frequency and components/band changed. Thus, subtle changes in the pattern of distribution of values for these parameters appeared to alter their relative salience. The results are consistent with predictions that parameters distributed categorically are more salient than noncategorically distributed parameters (e.g., Nosofsky, 1986; Sjoberg and Thorslund, 1979).

The three most significant acoustic parameters in this study are not unlike those reported by Allen and Bond (1997) who also used tones and tonal complexes in their stimulus set. The parameters that appeared most important to the ratings in that study included a measure of relative periodicity (equivalent to the number of components/band in the current study), the number of peaks in the amplitude spectra (equivalent to the number of frequency bands), and frequency. However, the relative weighting of these three parameters differed in the two studies. In the current study frequency proved to be most important but Allen and Bond found frequency was only the third most significant contributor to the ratings, correlating with dimension 3 in the fitted solution and accounting for a relatively small proportion of the variance in the ratings. Allen and Bond found that the number of peaks in the spectrum played a dominant role in the ratings but it was a much less significant contributor to the ratings in the current study. These differences across studies may be explained based upon the context in which the stimuli are presented and the pattern of distribution in values. The current study contained no noise stimuli thus the number of components/band covered a much smaller range than in the Allen and Bond study, possibly reducing its importance in the ratings. The lesser role played by frequency in the Allen and Bond study may be attributable to its relatively continuous distribution in that study and the fact that many stimuli had both high and low frequency components. The inability to use frequency to organize the stimuli could have reduced its importance.

Tversky (1977) suggests that a person will attempt to impose a categorical structure upon a stimulus set and make similarity judgments based upon those imposed categories. Stimulus properties that facilitate the formation of a cognitive structure will be more heavily weighted than other properties. The relatively categorical distribution of stimuli throughout the derived spaces of this and previous studies (e.g., Allen and Bond, 1997; Christensen and Humes, 1997) and the relative weighting of stimulus properties are consistent with this theory.

In a slightly different way of conceptualizing the formation of categories and similarity ratings, Ashby and Perrin (1988) suggest that, rather than the observer/listener imposing a categorical structure on the stimuli, the categories most likely arise from the stimuli themselves. According to their theory, each stimulus evokes a multidimensional perceptual response with which is associated a mean and variance. The perceptual response is a function of the integration of multiple parameter values (each of which is subject to the limitations of peripheral resolution and the process of integration by which multiple parameters are combined to form a perceptual unit). Perceived similarity in stimuli will vary with the degree of overlap in the multidimensional perceptual representations of the stimuli. Therefore, when stimulus parameter values are distributed throughout a stimulus set in a categorical manner, the perceptual responses elicited by the stimuli will show a smaller degree of overlap, potentially facilitating the formation of classification boundaries in the regions where the perceptual responses do not overlap. In the current study, the similarity judgments made by the listeners tended to reflect the underlying stimulus structure. For example, low frequency stimuli tended to be separated in the

space from high frequency stimuli and when frequency was distributed such that there was a larger break between the low and high frequency distributions, the positioning of stimuli within the multidimensional space reflected this. The tendency for stimuli to be organized in clusters within the multidimensional space, as was often observed in this and previous studies, may therefore reflect the underlying perceptual responses evoked by the stimuli.

In general, the findings from the current study suggest that the context in which stimuli are present, may affect the perceived salience of acoustic parameters as derived from similarity ratings. While previous studies have suggested that changes in the range over which a parameter can vary may affect salience, this study further suggests that subtle changes in the distribution of values within a fixed range may also affect salience as determined from the correlation between stimulus parameters and the dimensions of a derived multidimensional space.

V. SUMMARY AND CONCLUSIONS

Listeners' similarity ratings of multidimensional, unfamiliar sounds were evaluated for four sets of stimuli in a between subjects design. The acoustic parameters were very similar within each of the four sets but the distribution of values in two parameters was manipulated such that they varied either categorically or continuously. Results suggested that both temporal and spectral properties of stimuli were encoded in listeners' similarity ratings, including mean frequency, the number of components in each critical band (corresponding to a roughness percept), and the number of bands from which components were drawn (corresponding to the number of peaks in the amplitude spectra). When two stimulus properties (mean frequency and the number of components) were assigned values that were distributed categorically, their salience increased.

ACKNOWLEDGMENTS

The authors thank the listeners who participated in these studies and the Natural Sciences and Engineering Research Council of Canada and Unitron for financial support.

[1]It was theorized that listeners would be able to discriminate between stimuli containing one, two, three, or more components but that discrimination between stimuli containing four to nine components would be more difficult. When only a few tonal components lie within a critical band, they will produce a modulation in the temporal envelope that can be perceived as a roughness cue. The difference between one, two, or three tones should be perceptible as a difference in this cue. However, when many components lie within a critical band (e.g., 4–9), discrimination of differences in roughness should be less perceptible.

Allen, P., and Bond, C. A. (**1997**). ''Multidimensional scaling of complex sounds by school-aged children and adults,'' J. Acoust. Soc. Am. **102**, 2255–2263.

ANSI (**1989**). ANSI S3.6, 1989, ''Specifications for audiometers'' (American National Standards Institute, New York).

Ashby, F. G., and Perrin, N. A. (**1988**). ''Toward a unified theory of similarity and recognition,'' Psychol. Rev. **95**, 124–150.

Carroll, J. D., and Chang, J. J. (**1970**). ''Analysis of individual differences in multidimensional scaling via a n-way generalization of ''Eckart-Young'' decomposition,'' Psychometrika **35**, 283–319.

Christensen, L. A., and Humes, L. E. (**1996**). ''Identification of multidimensional complex sounds having parallel dimension structure,'' J. Acoust. Soc. Am. **99**, 2307–2315.

Christensen, L. A., and Humes, L. E. (**1997**). ''Identification of multidimensional stimuli containing speech cues and the effects of training,'' J. Acoust. Soc. Am. **102**, 2297–3210.

Fox, R. A. (**1983**). ''Perceptual structure of monphthongs and diphthongs in English,'' Lang Speech **26**, 21–47.

Nosofsky, R. M. (**1986**). ''Attention, similarity, and the identification–categorization relationship,'' J. Exp. Psychol. Gen. **115**, 39–57.

Scharf, B. (**1970**). ''Critical bands,'' in *Foundations of Modern Auditory Theory, Vol. 2*, edited by J. V. Tobias (Academic, New York).

Sjoberg, L., and Thorslund, C. (**1979**). ''A classificatory theory of similarity,'' Psychol. Res. **40**, 223–247.

Torgerson, W. S. (**1965**). ''Multidimensional scaling of similarity,'' Psychometrika **30**, 379–393.

Tversky, A. (**1977**). ''Features of similarity,'' Psychol. Rev. **84**, 328–352.

Spectral integration in bands of modulated or unmodulated noise[a)]

Sid P. Bacon,[b)] Nicolas Grimault, and Jungmee Lee
Psychoacoustics Laboratory, Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-1908

(Received 30 July 2001; revised 29 March 2002; accepted 8 April 2002)

Spectral integration was measured for pure-tone signals masked by unmodulated or modulated noise bands centered at the signal frequencies. The bands were typically 100 Hz wide, and when modulated, they were sinusoidally amplitude modulated at a rate of 8 Hz and a depth of 100%. In experiment 1, thresholds were first measured for each individual pure tone of a triplet in the presence of its respective masker band, and then for those three tones added together at their respective threshold levels, masked by their respective masker bands. Four sets of triplets were used: 250, 1000, 4000 Hz; 354, 1000, 2828 Hz; 500, 1000, 2000 Hz; and 800, 1000, 1200 Hz. When the masker bands were unmodulated, the amount of spectral integration was about 2.4 dB for all triplets, consistent with the integration expected based on the multiband energy detector model. When the bands were modulated, the amount of integration depended upon the spacing between masker bands; for the two widest spacings, the integration was between about 0 and 3 dB, whereas for the two closest spacings, the integration was approximately 5 dB. Experiments 2 and 3 addressed the cause of this greater spectral integration in the presence of the modulated masker bands with closer spacing. The second experiment demonstrated that sensitivity (d') was proportional to signal power regardless of whether the background noise was modulated or not, and thus the greater integration in dB in the presence of the modulated noise bands could not be accounted for by shallower psychometric functions in those conditions. Instead, the third experiment showed that the greater integration was likely due to the fact that the masker bands were comodulated. In other words, it was probably due to cues related to comodulation masking release when all three bands (and signals) were present. [DOI: 10.1121/1.1482072]

PACS numbers: 43.66.Dc, 43.66.Ba, 43.66.Mk [MRL]

I. INTRODUCTION

A number of studies have demonstrated that the auditory system is able to sum or integrate information across a wide frequency range in order to enhance the detectability of broadband or multicomponent signals. For example, Schafer and Gales (1949) demonstrated that signals composed of multiple sinusoidal components were detectable when each component was set a few dB below its threshold in isolation. Others have evaluated various aspects of spectral integration (e.g., Green, 1958; Green *et al.*, 1959; Spiegel, 1979; Buus *et al.*, 1986). Most of the results can be reasonably well understood in terms of the multiband energy detector model (e.g., Green, 1958), which assumes that the overall sensitivity to a multicomponent signal (d'_n) is

$$d'_n = \sqrt{\sum_{i=1}^{n} d'^2_i},$$

where d'_i is the sensitivity to a single sinusoidal component, and n is the number of such components. If the sensitivity to each component is the same, the formula reduces to

$$d'_n = \sqrt{n}\, d'_i .$$

In other words, if the sensitivity to each component is the same, the multiband energy detector model predicts that the overall sensitivity should increase with the square root of the number of components ($\sqrt{n}$).

One way in which spectral integration is often studied is within a masking paradigm. Most often, the masker has been an unmodulated broadband noise (e.g., Schafer and Gales, 1949; Green, 1958; Buus *et al.*, 1986). Because background noises are often modulated to some degree, it is important to extend studies of spectral integration to include modulated maskers. Grose and Hall (1997) did this recently. Their masker was a composite of narrow-band noises, each 20 Hz wide and centered at one of their sinusoidal signal components. These bands thus had relatively slow, inherent modulations associated with them. The amount of spectral integration was similar to that obtained by others in unmodulated broadband noise maskers, suggesting that the inherent modulations did not affect spectral integration. That conclusion, however, must be tempered by the fact that the individual components were not necessarily equally detectable, but instead were presented at equal physical levels. A different amount of integration might have been obtained had the individual sinusoidal signal components been adjusted to be equally detectable, as is typically done in studies of spectral integration. The purpose of the present study was to measure

[a)]Portions of this research were presented at the 134th meeting of the Acoustical Society of America [S. P. Bacon, N. Grimault, and J. Lee, "Spectral integration and the detection of tones in modulated and unmodulated noise," J. Acoust. Soc. Am. **102**, 3160 (1997)].

[b)]Electronic mail: spb@asu.edu

spectral integration in the presence of modulated and unmodulated noise maskers under conditions where each sinusoidal component in the multicomponent signal was equally detectable, to determine if the amount of spectral integration depends upon whether the noise masker is modulated or not.

II. EXPERIMENT 1: SPECTRAL INTEGRATION IN BANDS OF NOISE

A. Apparatus and stimuli

The signal was either a pure tone or three pure tones added together. There were a total of four such triplets. Each triplet consisted of a ''low,'' ''mid,'' and ''high'' frequency; the mid frequency was 1000 Hz for each triplet. The frequencies for the four triplets were: 250, 1000, and 4000 Hz; 354, 1000, and 2828 Hz; 500, 1000, and 2000 Hz; and 800, 1000, and 1200 Hz, respectively. The signals were presented in corresponding bands of noise. Each band was 100 Hz wide and was centered arithmetically at one of the pure-tone frequencies. The bands were created by adding pure tones spaced at 2-Hz intervals; the tones were all equal in amplitude with random starting phases. Only one permutation was used for each masker band (i.e., these were ''frozen'' noise bands). Each band was either unmodulated or sinusoidally amplitude modulated at a rate of 8 Hz and a depth of 100%. (Although each masker band has inherent fluctuations, it is designated as unmodulated unless it has sinusoidal amplitude modulation applied to it.) When modulated, the sinusoidal modulator for each band had a starting phase of $-\pi/2$ radians, so that the modulated masker began at its minimum. Thus, with the composite signal, the three masker bands were modulated coherently. These conditions do not represent typical comodulation masking release (CMR) conditions, however, because a signal is added to each band, not just to one band.

The signal and masker were digitally generated and produced at a 20 000-Hz sampling rate using separate Tucker-Davis Technologies (TDT DA1) digital-to-analog converters (DACs). The output of each DAC was low-pass filtered at 8000 Hz (Kemo VBF 25.01; 135 dB/oct) and attenuated before being added together and presented to the right earpiece of a pair of TDH-49P headphones mounted in an MX/51 cushion. The duration of the signal and masker was 500 ms, including 20-ms $\cos^2$ onset and offset ramps. The level per component within each masker band was 40 dB SPL, and was fixed throughout a block of trials. The level of the signal was varied adaptively via a programmable attenuator (Wilsonics PATT).

B. Procedure

For any given triplet, thresholds were first measured for each individual pure tone in the presence of its respective masker band. Thresholds were then measured for the composite signal, which consisted of those three tones added together at their respective threshold levels, masked by their corresponding masker bands.[1] Thresholds were measured using an adaptive, two-interval, forced-choice (2IFC) procedure that tracked 79.4% on the psychometric function (Levitt, 1971). Each run consisted of a block of 60 trials. The initial step size of 5 dB was reduced to 2 dB after the first two reversals. The first two (or three, if the total number of reversals was odd) reversal points were discarded, and the threshold estimate for a given block was based on the average of the (individual or composite) signal level at the remaining reversal points. On a few occasions, runs were discarded when there were fewer than six reversals in the mean or when the standard deviation of a given threshold estimate was greater than 5 dB. Final threshold values for a particular condition are based on the average of at least three runs. If the standard deviation of that average was greater than 3 dB, an additional estimate was obtained and included in the average. This continued until the standard deviation was less than 3 dB or until six estimates were obtained. For the most part, the thresholds for each trio of pure tones and their corresponding composite signal were obtained in a single, 2-h session. Even though the 1000-Hz tone was common to all four triplets, the threshold for it was measured eight separate times (for each of the four triplets in both modulated and unmodulated noise).

C. Subjects

Five subjects (four females, one male) participated. They ranged in age from 22 to 24 years, and had thresholds at or below 15 dB HL (ANSI, 1996) at audiometric test frequencies from 500 to 8000 Hz. One was the second author, two were other laboratory personnel, and the remaining two were paid for their participation. All subjects except one had previous experience as listeners in psychoacoustic experiments.

D. Results and discussion

Most (157/160) of the individual thresholds had standard deviations less than 3 dB. Moreover, the results were consistent across subjects, and thus the group mean results are shown in Fig. 1. Each panel shows the results for a given set of frequencies; the thresholds are shown separately for the low (L), mid (M), and high (H) frequencies, as well as for the combined (C) condition, where all three frequencies are presented together at relative levels corresponding to their respective thresholds obtained alone. Recall that the threshold level for the combined condition is the threshold level of the most intense of the three components. The triangles represent the thresholds obtained with the unmodulated masker (UN) and the squares represent those obtained with the modulated masker [designated as COM, as opposed to, say, MOD, to be consistent with later designations where results with coherently (COM) and incoherently (INC) modulated maskers are compared]. As expected, the thresholds are lower in the presence of the modulated masker; when averaged across all conditions, this threshold difference or masking release caused by modulating the masker was about 11 dB. These results are thus consistent with those of previous investigators (Carlyon *et al.*, 1989; Bacon *et al.*, 1997; Bacon and Lee, 1997; Eddins, 2001), who have observed similarly sized effects with relatively narrow masker bands. They also extend those previous studies by demonstrating that a masking release also can be observed for signals consisting

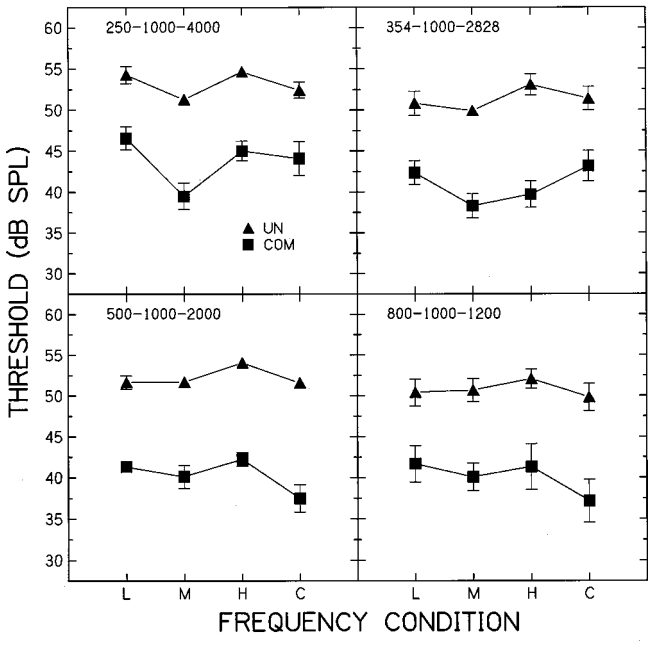


FIG. 1. Results from experiment 1, averaged across subjects. Each panel represents the results from a different triplet, as indicated. Thresholds are shown separately for the low (L), mid (M), and high (H) frequencies, as well as for the combined (C) condition, where all three signals are presented together. The error bars represent ±1 standard error (s.e.). The masker bands were either unmodulated (triangles) or modulated coherently at a rate of 8 Hz and a depth of 100% (squares).

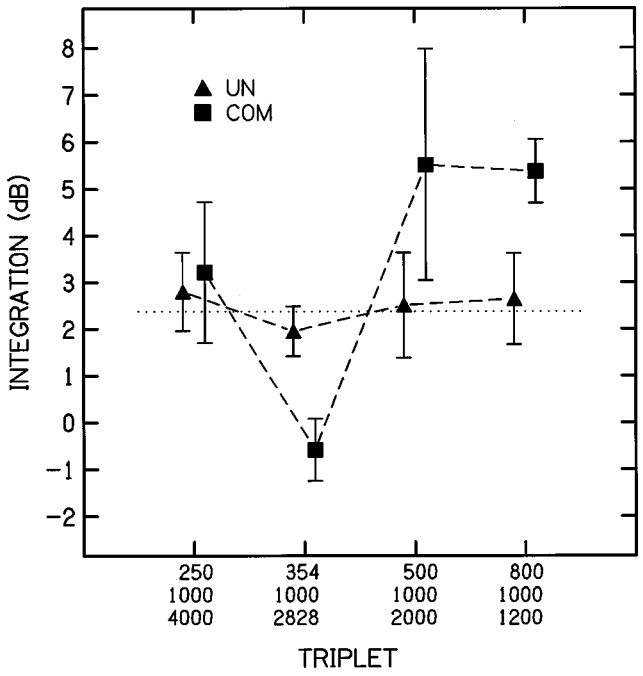


FIG. 2. The amount of spectral integration averaged across subjects from experiment 1 for the four different triplets. The amount of integration is defined as the difference between the highest threshold in the three single-tone conditions and the threshold in the combined condition. The masker bands were either unmodulated (triangles) or amplitude modulated coherently (squares). The error bars represent ±1 s.e. The dotted line corresponds to the 2.4 dB of integration that is expected based on the multiband energy detector model (see the text for details).

of multiple pure tones, and that the size of that release is similar to that obtained with each of the individual pure tones. As can be seen in Fig. 1, the size of the masking release is more or less independent of signal frequency. This, too, is consistent with previous results with narrow-band noise maskers, although with broadband maskers the masking release tends to increase with increasing signal frequency (Bacon *et al.*, 1997; Bacon and Lee, 1997; Eddins, 2001).

Of particular interest is whether the amount of spectral integration obtained in the presence of modulated masker bands differs from that obtained in the presence of unmodulated masker bands. Spectral integration is evidenced by a threshold in the combined condition that is lower than the highest threshold in the three corresponding single-tone conditions, because the signal level in the combined condition is the level of the component that had the highest threshold when measured in isolation. For example, consider the triangles in the upper-left panel. The highest threshold among the single-tone conditions is 54.7 dB SPL, and this is for the 4000-Hz tone (condition H). In the combined condition, the level of that 4000-Hz tone at threshold is 52.4 dB SPL. This indicates that, when combined, each sinusoidal component was 2.3 dB lower at threshold than it was when presented alone. In Fig. 2, we have plotted the spectral integration or improvement in threshold for all four triplets in both masking conditions (triangles: unmodulated; squares: modulated). Note that, for this figure, we have calculated the integration for each individual subject, and plotted the average integration across subjects. These values are nearly identical to the ones obtained by calculating the integration based on the average thresholds shown in Fig. 1. In addition to the obtained values, the amount of integration expected based on the multiband energy detector model is shown as a dotted line; in this case, it is assumed that d' is proportional to signal intensity (Egan *et al.*, 1969), and thus the model prediction in terms of dB is $10 \log \sqrt{n}$. In the presence of the unmodulated masker, the average amount of integration is essentially equal to the 2.4 dB that is predicted from the energy detector model. These results thus confirm previous results, obtained in either a broadband noise (Schafer and Gales, 1949; Green, 1958; Green *et al.*, 1959; Buus *et al.*, 1986) or a set of narrow-band noises (Grose and Hall, 1997). For example, Buus *et al.* (1986) found that the threshold for an 18-tone complex was 6.2 dB lower than the average threshold for three of those tones presented in isolation ($10 \log \sqrt{18} = 6.3$ dB), and Grose and Hall (1997) observed the expected 4.5-dB decrease in threshold when going from a signal with one to a signal with eight components. It is interesting to note that the $\sqrt{n}$ expected improvements were observed in these two studies despite the fact that the individual components were not necessarily equally detectable, but instead were presented at equal physical levels. Buus *et al.* used a uniformly masking noise, which tends to yield masked thresholds that are independent of signal frequency. And although this was generally true for the three (of the 18) tones that they studied in isolation, it was not strictly true, as the thresholds for the three tones in isolation sometimes differed by as much as 3 dB (see their Fig. 1). Grose and Hall, on the other hand, used 20-Hz-wide bands of noise as maskers (each with a spectrum level of 35 dB). They did not measure the threshold for each of the eight tones in isolation

(only the threshold for one), but based on our own threshold measurements, it seems highly unlikely that their tones were equally detectable.

In the presence of the modulated masker bands, the amount of spectral integration depended upon the triplets, and was close to the expected amount of 2.4 dB in only one instance, namely the one where the components were most widely spaced (250, 1000, and 4000 Hz; far left). Interestingly, in one condition (354, 1000, and 2828 Hz), there was essentially no spectral integration. This finding was consistent across all five subjects. It is unclear why there would be no integration in this case. Although this is the only triplet without a harmonic relationship, that fact did not affect the integration when the masker bands were unmodulated; it is unclear why the lack of harmonicity would have an affect on spectral integration only when the masker bands were modulated.[2] Finally, in the remaining two conditions—where the components were closest together—the improvement was about 5 dB. Thus, at least when the masker bands (and their respective signals) are relatively close together, the amount of spectral integration can be greater than that expected based on the multiband energy detector model. We consider two possible explanations for this finding in the remaining two experiments.

III. EXPERIMENT 2: PSYCHOMETRIC FUNCTIONS

A. Rationale and conditions

One possible explanation for the greater spectral integration in some of the modulated conditions is related to the underlying psychometric functions. The energy detector model assumes that the integration ($\sqrt{n}$ improvement) is in terms of d'. As noted above, the $10 \log \sqrt{n}$ improvement (in dB) assumes that d' is proportional to signal intensity. Although this is generally the case (e.g., Egan *et al.*, 1969), most psychometric functions describing the detection of a tone have been obtained in quiet or in the presence of an unmodulated broadband noise. If the slope of the psychometric function were more shallow in the presence of a modulated masker, then a given integration in terms of d' would correspond to a larger amount of integration in terms of an improvement in threshold in dB. Thus, the purpose of this experiment was to compare the slopes of the psychometric functions for tones in the presence of unmodulated bands of noise with those obtained in the presence of modulated bands of noise. Only one set of triplets was used: 500, 1000, and 2000 Hz. This set yielded the largest integration in dB in experiment 1. These signals and maskers were identical to those in experiment 1, with the following exception. The masker bands for one subject (the second author, NG) were one-ERB (equivalent rectangular bandwidth) wide (Glasberg and Moore, 1990), rather than 100 Hz wide; thus, their bandwidths were 78, 132, and 240 Hz for the 500-, 1000-, and 2000-Hz tones. The data from this subject were collected prior to our decision to use 100-Hz-wide bands. It is unlikely that the results from the present experiment were influenced by these differences in bandwidths: NG completed the conditions of experiment 1 with this triplet and the ERB maskers, and those results were essentially identical to those obtained with the 100-Hz-wide masker bands.

For each signal condition (each single tone and all three combined), four signal levels were chosen (five in one case for NG) in order to straddle the threshold level obtained in experiment 1. Each of the points on the psychometric function was based on 300 trials (5 blocks of 60).

B. Subjects

Three of the five subjects (one male and two females) from experiment 1 participated; two were paid for their participation.

C. Results and discussion

The psychometric functions for the individual subjects are shown in separate panels of Fig. 3. The results are plotted in terms of d' on a log scale versus signal level in dB (in the combined case, the signal level is the level of the most intense single component in the three-tone signal). The unfilled symbols represent the results with unmodulated masker bands, whereas the filled symbols represent the results with the modulated masker bands. Averaged across subjects, the amount of spectral integration in dB (at $d'=1$) was 3.2 dB in the unmodulated condition and 5.6 dB in the modulated condition. These results are thus consistent with those of experiment 1.

Each psychometric function was fitted with a straight line representing the best-fitting linear regression analysis. In one instance (JL, 2000 Hz, unmodulated), one of the points was not included in the fit (nor plotted) because performance was just below chance, leading to a $d'<0$, which could not be plotted on this y axis. The fits were generally good to excellent, with r^2 values ranging from 0.75 to 0.99 (mean of 0.94). When averaged across subjects and conditions, the slope was 0.117 in the unmodulated conditions and 0.122 in the modulated conditions. Note that these slopes are in terms of signal level in dB; they are a factor of 10 less than those if $\log d'$ were plotted as a function of log signal power. Those slopes, 1.17 and 1.22, are close to the expected slopes of about 1.0 for detection of tones in quiet or noise (Egan *et al.*, 1969). The similarity in slopes between the modulated and unmodulated conditions indicates that the greater integration in the modulated condition is not due to shallower psychometric functions in the presence of modulated masker bands.

IV. EXPERIMENT 3: COMODULATED VERSUS INCOHERENTLY MODULATED BANDS

A. Rationale and conditions

Because coherent amplitude modulation was applied to the masker bands in the first two experiments, another possible explanation for the greater spectral integration in some of the modulated conditions is that cues related to CMR resulted in a relatively large threshold improvement when all three bands (and signals) were present. Consistent with this possibility is the fact that the especially large spectral integration only occurred when the masker bands were relatively close to one another (Fig. 2), as CMR tends to decrease as

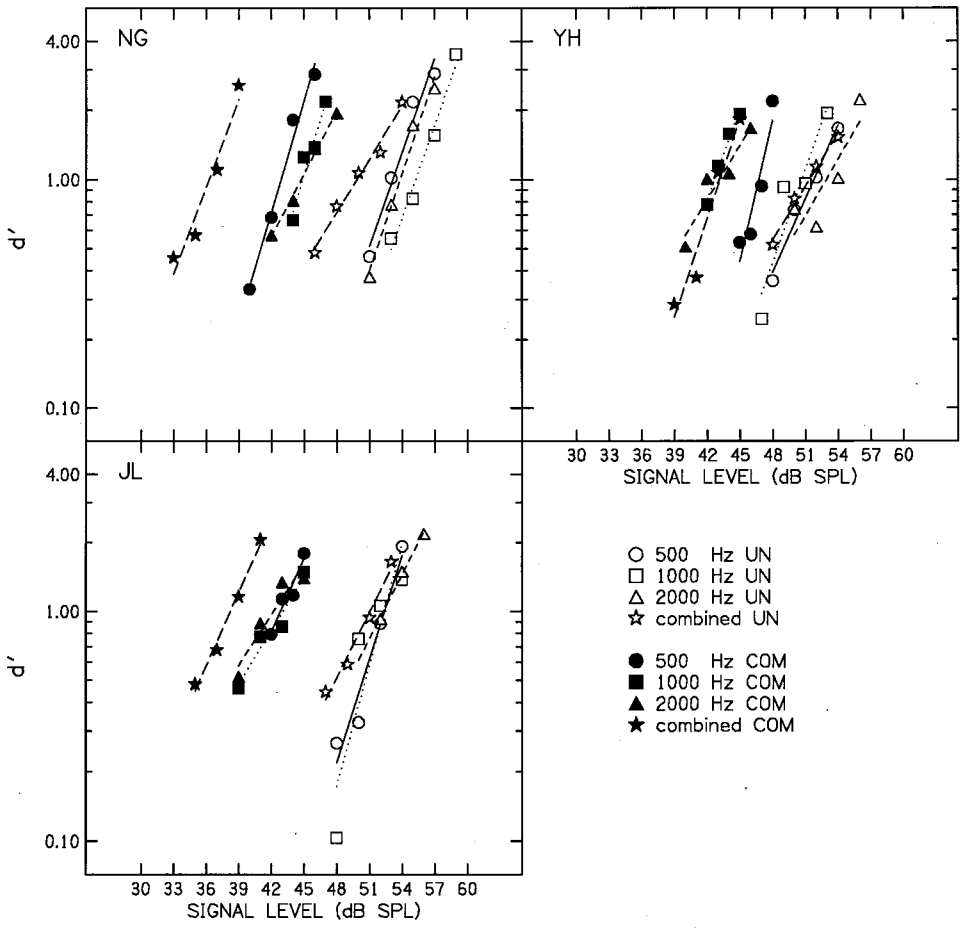


FIG. 3. Individual results from experiment 2. Psychometric functions are shown for the detection of each individual tone in a triplet (500, 1000, and 2000 Hz), as well as for the condition where all tones are presented together. The masker bands were either unmodulated (unfilled symbols) or amplitude modulated coherently (filled symbols). Each function was fitted with a straight line based on linear regression analysis.

the distance between masking bands increases (Cohen and Schubert, 1987; Hall, 1987; Schooneveldt and Moore, 1987). The masking release observed in CMR experiments presumably results from a difference in the on-frequency and flanking band envelopes caused by the addition of the signal to the on-frequency masking band. In our experiments we add a signal to each of the bands, and thus our across-band differences will be much smaller than those in a typical CMR experiment. It is worth emphasizing, however, that the across-band differences are not necessarily eliminated, because the signal levels are typically different in each band. Thus, cues that result in a masking release in a typical CMR experiment could be responsible for the especially large spectral integration observed in the previous two experiments where the masker bands were relatively close to one another. The purpose of this experiment was to address that possibility. Only one set of triplets was used: 500, 1000, and 2000 Hz. This set yielded the largest integration in dB in experiment 1. The signals and maskers were identical to those in experiment 1; in addition, a set of incoherently modulated masker bands was used. The modulator phase for the low, mid, and high bands was $-\pi/2$, 0, and $\pi/2$ radians, respectively.

B. Subjects

Four female subjects participated; they ranged in age from 22 to 24 years. One had participated in experiment 1, and her data from the unmodulated and comodulated conditions are repeated here. All subjects had previous experience as listeners in psychoacoustic experiments, audiometric thresholds less than 15 dB HL (ANSI, 1996), and all were associated with the laboratory.

C. Results and discussion

All but one (47/48) of the individual thresholds had standard deviations less than 3 dB, and the pattern of results was similar across subjects. Thus, the group mean results are shown in Fig. 4, where the amount of spectral integration is shown for the unmodulated and two modulated conditions. The expected integration (2.4 dB) based on the multiband energy detector model is shown as an asterisk (far left). The amount of integration in the unmodulated condition (UN) was 1.5 dB. When the bands were comodulated (COM), the integration was about 6 dB, similar to that observed in experiments 1 and 2. When the bands were incoherently modulated (INC), however, the improvement was only about 2 dB, in line with the $\sqrt{n}$ improvement. These results suggest that the greater spectral integration observed in the presence of modulated noise bands in experiments 1 and 2 is due to the fact that the bands were modulated coherently. This, in turn, suggests that cues related to CMR were responsible for the greater spectral integration with modulated noise bands. Thus, it appears that CMR can play a role in the present experiments, despite the fact that we did not employ a typical CMR paradigm.

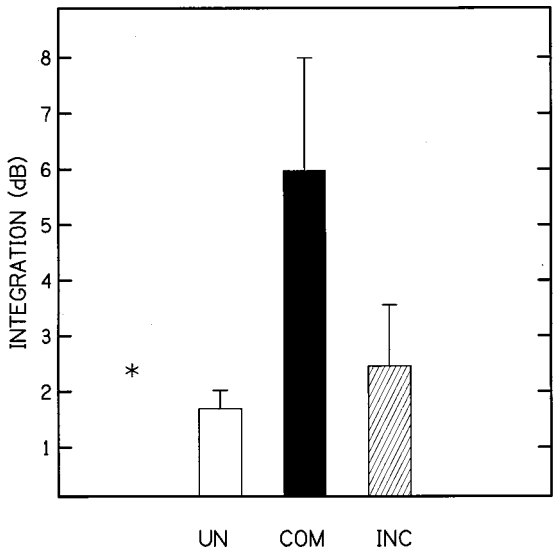


FIG. 4. The amount of spectral integration (+1 s.e.) for the triplet 500, 1000, and 2000 Hz averaged across subjects from experiment 3. The masker bands were unmodulated (UN), comodulated (COM), or incoherently modulated (INC). The asterisk indicates the 2.4 dB of integration that is expected based on the multiband energy detector model.

V. GENERAL DISCUSSION

The amount of spectral integration in the presence of the unmodulated masker bands was consistent with the amount predicted by the multiband energy detector model. These results are thus similar to previous results on spectral integration of pure tones in the presence of broadband noise maskers (Schafer and Gales, 1949; Green, 1958; Buus *et al.*, 1986). Furthermore, they confirm the results of Grose and Hall (1997), who measured spectral integration in the presence of multiple, 20-Hz-wide maskers. The similarity between our results and those of Grose and Hall suggests that their results were not influenced much by the fact that their individual pure-tone signals were not necessarily equally detectable. Moreover, because our unmodulated masker bands and their masker bands contained inherent fluctuations, the results from both studies indicate that irregular, inherent fluctuations in maskers do not influence the amount of spectral integration for multicomponent signals, at least when those fluctuations are incoherent or independent. No one has examined whether relatively slow, *inherent* fluctuations in masker bands influence spectral integration when they are coherent. The results of the present study suggest that they might, given that when sinusoidal modulation was imposed coherently on each masker band, the amount of integration was sometimes larger than that predicted by the multiband energy detector model. In particular, experiment 1 showed that the amount of spectral integration for a triplet of pure tones was about 3 dB greater than predicted for the two triplets with the narrowest frequency separations.

The decline in the amount of integration when going from the two triplets with the narrowest spacing to the two with the widest spacing could represent a general spatial limit to spectral integration. In that sense, it is not inconsistent with some results on spectral integration in gap detection. Grose (1991) found that the improvement in gap detection threshold when going from one to multiple narrow-band markers was greater when those markers were relatively closely spaced than when they were more widely spaced. However, the effect of spacing in the present study depended upon whether the masker bands were modulated or not: there was no effect of spacing when the masker bands were unmodulated. Furthermore, the effect of spacing with the modulated masker bands exists not because the amount of integration was especially small at wide separations, but because it was especially large at narrower separations. This implies that there is something unique about the auditory processing underlying spectral integration in the presence of comodulated masker bands that are spaced relatively closely together (i.e., within an octave of one another).

We argued previously that the greater spectral integration in the presence of the more closely spaced masker bands was due to the same auditory processing that underlies CMR. Consistent with that possibility was the finding that spectral integration decreased—and was consistent with the amount expected based on the energy detector model—when the masker bands were modulated incoherently (Fig. 4). Furthermore, the effect of spacing observed here is consistent with the effect of spacing observed in CMR experiments (Cohen and Schubert, 1987; Hall, 1987; Schooneveldt and Moore, 1987). This suggests that there may be two different types of processing affecting the amount of spectral integration in the presence of relatively closely spaced, comodulated masker bands. One is the processing that underlies spectral integration, and the other is the processing that underlies CMR. If the effects of the two were additive, we could estimate how much of the measured integration was due to each type of processing. As noted above, the amount expected based on spectral integration *per se* is 2.4 dB. Because the maximum amount of measured integration was about 5–6 dB, this suggests that about 3 dB of the measured integration is due to processes underlying CMR. Because the stimulus situation in the present study, where a tone was added to each masker band, is so different from that found in typical CMR studies, it is difficult to determine whether this estimated value is consistent with the literature on CMR. It is clear that one would not want to compare this estimated amount of CMR with the amount of CMR obtained under a more typical CMR condition in which the masker bands were all modulated at a depth of 100% and the signal was added to the middle band. The amount of CMR obtained in this condition would almost certainly be larger than that obtained under stimulus situations comparable to those in the present study. Indeed, our stimulus situation is more like that of detecting a single tone in the presence of flanking bands with a reduced modulation depth, because each signal will reduce the modulation depth of its respective masker band. Thus, the most relevant comparisons are with studies in which the masker and flanking bands were modulated at depths different from one another, and all less than 100% (Grose and Hall, 1989; Fantini, 1991; Hicks and Bacon, 1995). Even then, the comparisons must be made with caution, given that the stimulus situations in those studies were still quite different from those in the present study (some of the differences include the number, width, and spacing of the masker bands). Of

those studies, the one by Fantini is perhaps most similar to the present study; she typically observed small amounts of CMR, often 3 dB or less, consistent with the estimated amount of CMR in the present study.

If the larger amounts of integration for relatively closely spaced signals and comodulated maskers is due to processing underlying CMR, it is of interest to consider specifically what that processing might be. As discussed by others (e.g., Hall and Grose, 1988), there are multiple cues or different types of processing that can lead to CMR. These include, for example, across-frequency differences in envelope correlation, across-frequency differences in envelope amplitude, and a cued listening in the dips of the masker envelope. Presumably, any of these or other cues could have played a role in the present study. Because the stimuli were not designed to distinguish among these cues, the results do not implicate one model or type of processing over another.

VI. SUMMARY AND CONCLUSIONS

The purpose of the present study was to determine if the amount of spectral integration with multicomponent signals is influenced by whether the background noise is modulated or not. Three experiments were conducted, wherein spectral integration was measured by comparing the detectability of a signal consisting of one of three pure tones with that of a signal consisting of all three tones presented together. These pure tones were masked by narrow bands of noise (typically 100 Hz wide) centered at the pure-tone frequencies; the bands were unmodulated or sinusoidally amplitude modulated at a rate of 8 Hz and a depth of 100%. The results can be summarized as follows.

(1) In the first experiment, spectral integration was measured for three triplets. When the masker bands were unmodulated, the amount of spectral integration was consistent with the amount of integration predicted by the multiband energy detector model. When the bands were amplitude modulated, the amount of integration depended upon the frequency separation between the tones (and their corresponding masker bands): for the two triplets with the narrowest frequency separations, the amount of spectral integration was about 5–6 dB.

(2) In the second experiment, psychometric functions were measured for the single- and combined-tone signals of one of the triplets with a relatively narrow frequency separation (500, 1000, and 2000 Hz). The slopes of the psychometric functions were close to 1.0 for both the unmodulated and modulated masker bands, indicating that differences in integration in dB were not due to differences in the slopes of the underlying psychometric functions.

(3) Finally, in the third experiment, spectral integration was measured as in experiment 1 (with only one triplet: 500, 1000, and 2000 Hz), but in this case with two different sets of modulated bands: one set was comodulated (as in experiments 1 and 2) and the other was incoherently modulated. The amount of integration was similar when the bands were unmodulated or incoherently modulated (1.5–2.0 dB), but was considerably larger (6.0 dB) when the bands were comodulated. These results suggest that the greater integration in the comodulated condition is due to cues related to CMR.

The results of the present study thus indicate that spectral integration is probably unaffected by the type of background noise, as long as cues related to CMR do not exist. Of course, such cues may exist in typical, real-world listening situations, thus enhancing (at least somewhat) the auditory system's ability to integrate information across frequency regions in everyday environments.

ACKNOWLEDGMENTS

This research was supported by NIDCD Grant No. DC01376. We thank Tammo Houtgast and the other reviewer for their comments on a previous version of this manuscript.

[1]For example, if the thresholds for the individual low, mid, and high frequencies were 50, 40, and 45 dB SPL, respectively, then the composite signal would consist of those three tones at the *relative* levels of 0, −10, and −5 dB. The level of the composite signal is designated as the level of the most intense tone. Thus, if the composite signal level were 50 dB at threshold, there would be no spectral integration. If, on the other hand, the composite signal level at threshold were 45 dB, then there would be 5 dB of spectral integration, as the levels would be 45, 35, and 40 dB SPL.

[2]Supplementary data collected from one of the original subjects (the second author) and an additional subject with considerable experience as a listener in psychoacoustic tasks suggest that integration may indeed be affected when inharmonic masker bands are modulated. In one set of conditions, the 100-Hz-wide masker bands were centered at 631, 1000, and 1585 Hz. When the bands were unmodulated, the average integration was about 3 dB; when the bands were modulated, it was about 0 dB. In this case, the modulation was incoherent (the modulator phase for the low, mid, and high bands was $-\pi/2$, 0, and $\pi/2$ radians, respectively). In another set of conditions, masker bands were centered at 500, 1000, and 2000 Hz. In this case, the average amount of integration was close to 2 dB for both the unmodulated and incoherently modulated masker bands. We have no explanation for the rather consistent lack of integration when inharmonic masker bands are modulated, although these supplementary data suggest that it is not idiosyncratic to the frozen noise samples used in experiment 1.

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

Bacon, S. P., and Lee, J. (**1997**). "The modulated-unmodulated difference: Effects of signal frequency and masker modulation depth," J. Acoust. Soc. Am. **101**, 3617–3624.

Bacon, S. P., Lee, J., Peterson, D. N., and Rainey, D. (**1997**). "Masking by modulated and unmodulated noise: Effect of bandwidth, modulation rate, signal frequency, and masker level," J. Acoust. Soc. Am. **101**, 1600–1610.

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.

Carlyon, R. P., Buus, S., and Florentine, M. (**1989**). "Comodulation masking release for three types of modulator as a function of modulation rate," Hear. Res. **42**, 37–46.

Cohen, M. F., and Schubert, E. D. (**1987**). "Influence of place synchrony on detection of a sinusoid," J. Acoust. Soc. Am. **81**, 452–458.

Eddins, D. A. (**2001**). "Measurement of auditory temporal processing using modified masking period patterns," J. Acoust. Soc. Am. **109**, 1550–1558.

Egan, J. P., Lindner, W. A., and McFadden, D. (**1969**). "Masking-level differences and the form of the psychometric function," Percept. Psychophys. **6**, 209–215.

Fantini, D. A. (**1991**). "The processing of envelope information in comodulation masking release (CMR) and envelope discrimination," J. Acoust. Soc. Am. **90**, 1876–1888.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Green, D. M. (**1958**). "Detection of multiple component signals in noise," J. Acoust. Soc. Am. **30**, 904–911.

Green, D. M., McKey, M. J., and Licklider, J. C. R. (**1959**). ''Detection of a pulsed sinusoid in noise as a function of frequency,'' J. Acoust. Soc. Am. **31**, 1446–1452.

Grose, J. H. (**1991**). ''Gap detection in multiple narrow bands of noise as a function of spectral configuration,'' J. Acoust. Soc. Am. **90**, 3061–3068.

Grose, J. H., and Hall, J. W. (**1989**). ''Comodulation masking release using SAM tonal complex maskers: Effects of modulation depth and signal position,'' J. Acoust. Soc. Am. **85**, 1276–1284.

Grose, J. H., and Hall, J. W. (**1997**). ''Multiband detection of energy fluctuations,'' J. Acoust. Soc. Am. **102**, 1088–1096.

Hall, J. W. (**1987**). ''Experiments on comodulation masking release,'' in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ).

Hall, J. W., and Grose, J. H. (**1988**). ''Comodulation masking release: Evidence for multiple cues,'' J. Acoust. Soc. Am. **84**, 1669–1675.

Hicks, M. L., and Bacon, S. P. (**1995**). ''Some factors influencing comodulation masking release and across-channel masking,'' J. Acoust. Soc. Am. **98**, 2504–2514.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Schafer, T. H., and Gales, R. S. (**1949**). ''Auditory masking of multiple tones by random noise,'' J. Acoust. Soc. Am. **21**, 392–398.

Schooneveldt, G. P., and Moore, B. C. J. (**1987**). ''Comodulation masking release (CMR): Effects of signal frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band,'' J. Acoust. Soc. Am. **82**, 1944–1956.

Spiegel, M. F. (**1979**). ''The range of spectral integration,'' J. Acoust. Soc. Am. **66**, 1356–1363.

Effects of signal delay on auditory filter shapes derived from psychophysical tuning curves and notched-noise data obtained in simultaneous masking[a)]

Sid P. Bacon,[b)] Jennifer L. Repovsch-Duffey, and Li Liu
Psychoacoustics Laboratory, Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-1908

(Received 29 May 2001; revised 28 March 2002; accepted 12 April 2002)

Psychophysical tuning curves (PTCs) measured in simultaneous masking usually sharpen as a short duration signal is moved from the onset to the temporal center of a longer duration masker. Filter shapes derived from notched-noise maskers have not consistently shown this effect. One possible explanation for this difference is that the signal level is fixed in the PTC paradigm, whereas the masker level is usually fixed in the notched-noise paradigm. In the present study, the signal level was fixed at 10 dB SL in both paradigms. The signal was 20 ms in duration, and presented at the onset or temporal center of the 400-ms masker. The masker was a pure tone presented in quiet (PTC) or in the presence of a pure-tone ''restrictor'' intended to limit off-frequency listening (PTCr), or it was a noise with a spectral notch placed symmetrically or asymmetrically about the 2-kHz signal frequency. Filter shapes were derived from the PTC, PTCr, and notched-noise data using the roex (p, w, t) model. The effects of signal delay and masking paradigm on filter bandwidth were analyzed with a two-factor repeated-measures ANOVA. There was a significant effect of signal delay (the filters sharpened with time) and masking paradigm (the filters derived from the notched-noise data were significantly wider than those derived from either of the PTC measurements, which did not differ from one another). Although the interaction between delay and paradigm was not significant, the filter derived from the notched-noise data sharpened more with time than did the other filters, and thus the bandwidth of the filters from the three paradigms were more similar at the longer delay than at the shorter delay. It is likely that the tuning-curve and notched-noise paradigms measure the same underlying filtering, but that various other factors contribute differentially to the derived filter shapes. [DOI: 10.1121/1.1485972]

PACS numbers: 43.66.Dc, 43.66.Mk, 43.66.Ba [MRL]

I. INTRODUCTION

Although the spectral and intensive properties of simultaneous masking have received the preponderance of attention in the literature, a considerable amount of research has focused on temporal properties, either by measuring the threshold for a short duration signal at different times within a longer duration masker, or by comparing the masking produced by a gated versus a continuous masker. Using tonal maskers, Bacon and Viemeister (1985a, b) showed that the largest temporal effect occurs when the masker frequency (f_m) is higher than the signal frequency (f_s). An interesting consequence of the fact that the temporal effect depends upon the $f_m - f_s$ relation is that psychophysical measures of frequency selectivity depend upon the temporal relation between masker and signal. For example, psychophysical tuning curves (PTCs) sharpen during the first 100–200 ms of masker stimulation, especially on the high-frequency side (Bacon and Viemeister, 1985a; Bacon and Moore, 1986; Kimberley *et al.*, 1989).

Although experiments with tonal maskers have consistently revealed a sharpening of psychophysical frequency selectivity with time, experiments with noise maskers have not. In particular, filter shapes derived from notched-noise masking data have not always shown a sharpening with time. Moore *et al.* (1987) found that the auditory filter shapes derived from their mean results across three subjects did not sharpen with signal delay. They concluded that auditory filtering is instantaneous and does not develop over time, and that the temporal effect shown with PTCs is due to aspects other than auditory filtering *per se*. Wright and Dai (1994), on the other hand, reported a narrowing of the derived auditory filter as they delayed their signal from masker onset. In addition, others (Carlyon, 1989; Hicks and Bacon, 1992) showed evidence, to varying degrees, of a temporal effect with notched-noise maskers, although the actual filter shapes were not derived in those studies.

It is unclear why there are inconsistencies with regard to the effect of time on measures of frequency selectivity obtained using tonal and noise maskers. It might be related to the fact that the filters based on notched-noise data have all been derived with the assumption that masker level determines the filter shape. According to Rosen and Baker (1994) and Rosen *et al.* (1998), it may be more appropriate to derive the auditory filter with the assumption that the signal level determines the shape. This can be accomplished either by

[a)]Portions of this research were presented at the 140th meeting of the Acoustical Society of America [J. L. Repovsch and S. P. Bacon, ''Psychophysical tuning curves and filter shapes sharpen over time,'' J. Acoust. Soc. Am. **108**, 2572(A) (2000)].

[b)]Electronic mail: spb@asu.edu

analyzing the results of experiments with a fixed masker level differently, or by performing the notched-noise experiments with a fixed signal level, as was done, for example, by Glasberg *et al.* (1984a) and Lutfi and Patterson (1984). This latter approach, of course, is more similar to the way in which PTCs are measured.

Another factor that could contribute to the discrepancy in the literature regarding how PTCs and derived filter shapes change with time is individual differences, which can be sizable in experiments measuring temporal effects in simultaneous masking. Given these differences, it is important to measure the effects of time on both measures of frequency selectivity in the same group of subjects, something that has hitherto not been done.

The primary purpose of the present study was to measure the effects of signal delay on PTCs and filter shapes derived from notched-noise data in the same subjects using a fixed signal level. A secondary purpose was to compare the degree of tuning revealed by these two commonly used paradigms. There are apparently no published studies comparing PTCs and filter shapes in the same subjects, although both are used extensively as psychophysical measures of tuning, and their similarities and differences have been discussed (e.g., Glasberg *et al.*, 1984b). One important difference between the two paradigms is that off-frequency listening can markedly influence the shape of the PTC; off-frequency listening is less likely to occur in the notched-noise paradigm and, more importantly, it is taken into account by the algorithm that derives the underlying filter. As shown by Johnson-Davies and Patterson (1979; also see, for example, O'Loughlin and Moore, 1981), when off-frequency listening is limited by the use of a ''restrictor,'' the PTC becomes broader. Thus, comparisons of the tuning revealed by the PTC and notched-noise paradigms should take off-frequency listening into account. This was done in the present study in two ways. In one, the tuning curves were obtained without (PTC) and with (PTCr) a restrictor. In the other, filter shapes were derived from the PTC, PTCr, and notched-noise data using the same roex model (Patterson *et al.*, 1982). This latter approach has the advantage of accounting for off-frequency listening in the same way for both the PTC and notched-noise paradigms, and it allows a more complete and arguably a more fair and direct comparison between the two measures of frequency selectivity.

II. METHOD

A. Subjects

Four subjects participated, one male and three females. Their ages ranged from 23 to 35 years. All had absolute thresholds of 15 dB HL or better (ANSI, 1996) for octave frequencies from 250 to 8000 Hz. Subject S1 was the second author. The remaining subjects were paid an hourly wage for their participation.

B. Stimuli

The 2-kHz signal was 20 ms in duration with 10-ms $\cos^2$ rise/fall times. The masker was either a tone or a noise; it was 400 ms in duration with a 10-ms $\cos^2$ rise/fall time. The signal was located at the onset (0-ms delay) or at the temporal center (190-ms delay) of the masker.

The sinusoidal signal was digitally generated and produced at a 20-kHz sampling rate. The output of the digital-to-analog converter or DAC (TDT DA1) was low-pass filtered at 8 kHz (Kemo VBF 25.01; 135 dB/oct).

The tonal masker used in the PTC paradigm also was digitally generated, produced via the second channel of the DAC, and low-pass filtered at 8 kHz (Kemo VBF 25.01). The frequency of this masker was 1.00, 1.30, 1.50, 1.70, 1.90, 1.95, 2.0, 2.10, 2.20, 2.30, 2.50, 2.70, or 3.00 kHz. The starting phase of the masker was 90 degrees, while the starting phase of the signal was 0 degrees. A low-pass noise was sometimes presented to mask the cubic difference tone (CDT), $2f_m - f_s$. The noise, produced by a noise generator (GenRad 1381), was low-pass filtered (Kemo VBF 25.01) at a cutoff frequency of $2f_m - f_s$, and was present only when f_m was 1.30, 1.50, or 1.70 kHz. These are the conditions where the CDT was most likely to be detected (Greenwood, 1971). The low-pass noise was presented continuously at a spectrum level that was 30 dB below the level of the signal. This level should be more than sufficient to mask the CDT, given estimates of its level (Zwicker, 1979) and the appropriate critical-ratio data (Hawkins and Stevens, 1950).

The restrictor, when used, was a continuous tone (HP 8904A synthesizer) with a frequency of 1.8 kHz ($f_m > f_s$) or 2.2 kHz ($f_m < f_s$). Typically, only a subset of the masker frequencies used for the PTC paradigm was used for the PTCr paradigm, and the on-frequency condition ($f_m = 2.0$ kHz) was not run. The level of the restrictor was chosen to be the highest level that would not produce any masking of the 2.0-kHz signal. This level was 15 dB below the corresponding masker level on the PTC in the 190-ms-delay condition for all four subjects.

The notched-noise maskers were produced by multiplying each of two sinusoids (HP 8904A synthesizer) by a separate band of noise (GenRad 1381) that had been low-pass filtered at 0.40 kHz (Kemo VBF 25.01). The bandwidth of each multiplied band of noise was twice the cutoff frequency of the low-pass filter. The two independent bands of noise were added together and then led to a multiplier. The second input to this multiplier was a digital gating function (TDT DA1; 20-kHz sampling rate) that was low-pass filtered at 8 kHz (Kemo VBF 25.01); this was used to turn the notched-noise masker on and off. The resultant noise masker then consisted of two independent 800-Hz-wide bands placed symmetrically or asymmetrically about f_s. In the symmetric conditions, the closer edge of the noise bands was placed at 0.0 (no-notch condition), 0.1, 0.2, 0.3, 0.4, or 0.5 f_s. In the asymmetric condition, one edge was always 0.2 further away from f_s than the other edge. The edge of the closer band was at 0.1, 0.2, 0.3, or 0.4 f_s.

The stimuli were added together and delivered monaurally to the right ear via a TDH-49P headphone mounted in an MX/51 cushion. All timing, presentation of stimuli, and recording of responses was controlled by a personal computer (Compaq 486/66), and all parameters of the signal and masker were verified with a digital storage oscilloscope

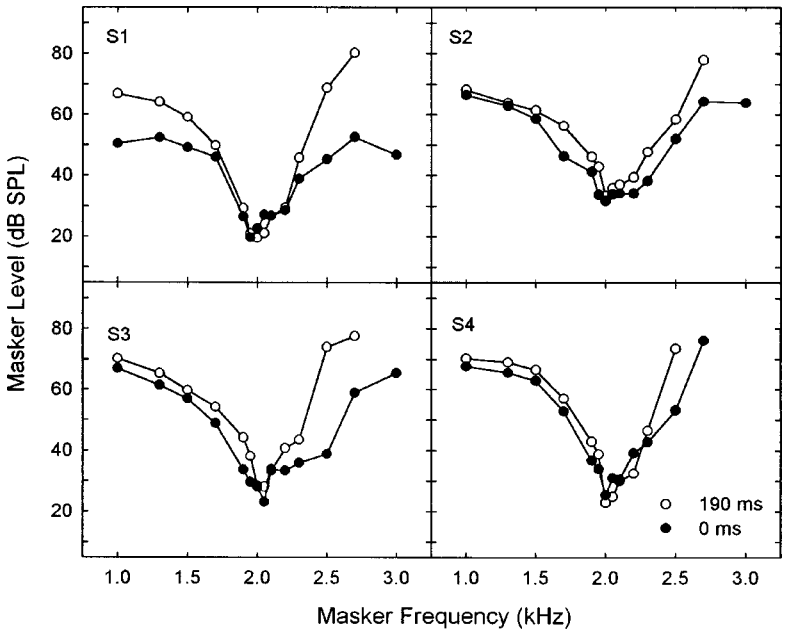


FIG. 1. PTCs for individual subjects for the 0-ms (filled circles) and 190-ms (unfilled circles) signal delays. The level of the signal was 10 dB SL, which corresponds to 27, 31, 30, and 28 dB SPL for subjects S1–S4, respectively.

(Tektronix 2230) and a spectrum analyzer (Hewlett-Packard 3561A).

C. Procedure

Each subject was tested individually in a double-walled sound booth. A two-interval forced-choice (2IFC) procedure was used, with an adaptive rule that tracked 79.4% correct (Levitt, 1971). Two subjects (S1, S3) were tested in the PTC paradigm first, and the other two were tested in the notched-noise paradigm first; the PTCr paradigm was always tested last. Within a paradigm, the masker conditions and signal delays were tested in a quasi-random fashion.

Subjects' answers were recorded from a response box. Lights on the response box were used to mark the observation intervals and to provide correct-answer feedback. Each run consisted of 60 trials. The signal was fixed in level at 10 dB SL, while the masker was varied adaptively. A run typically began with the masker 10–15 dB below masked threshold. The level of the masker increased after three correct responses, and decreased after one incorrect response. A reversal was defined as any point at which the direction changed from increasing to decreasing or vice versa. An initial step size of 5 dB was used for the first two reversals, followed by a step size of 2 dB for the remainder of the reversals. The first two (or three) reversals were discarded, and the threshold estimate for the run was then determined using the masker levels at the remaining even number of reversal points. A run was discarded if the standard deviation of the threshold estimate was greater than 5 dB, or if the number of reversals in the mean was less than six. A threshold for a given condition consisted of at least three runs obtained over at least 2 days. If the standard deviation of the mean of those three runs was greater than 3 dB, an additional threshold estimate was obtained and included in the average. This was done until the standard deviation was less than 3 dB, or until six total estimates were obtained. Approximately 88% of the thresholds had a standard deviation of less than 4 dB. Variability was greater in the onset condition, as discussed in more detail later.

Thresholds in quiet for the 20-ms signal were obtained using a similar 2IFC procedure. In this case, however, the level of the signal was varied adaptively. These thresholds, which are listed in the caption to Fig. 1, had standard deviations less than 3 dB.

III. RESULTS AND DISCUSSION

Figure 1 shows the PTC for each of the four subjects. Filled circles show the results for the signal located at the onset of the masker and unfilled circles show those for the signal at the temporal center of the masker. Although there was some variability among subjects in terms of the magnitude of the temporal effect, the PTC for the signal at the temporal center of the masker was sharper than the one for the signal at the onset of the masker for all subjects. This sharpening with time occurred on both the low- and high-frequency sides, but it tended to be greater on the high-frequency side. These effects are similar to those observed previously (Bacon and Viemeister, 1985a; Bacon and Moore, 1986; Kimberley *et al.*, 1989).

The unfilled circles in Fig. 2 show the PTCr for each subject. The top panels show the results for the 0-ms condition, whereas the bottom panels show the results for the 190-ms condition. The results from Fig. 1 are replotted as filled circles for comparison. The size of the effect of the restrictor varied across subjects, but there was a trend for the PTCr to be at least somewhat broader than the PTC, at both delays. The effect of the restrictor was much smaller than that observed by Johnson-Davies and Patterson (1979) and, to a lesser extent, O'Loughlin and Moore (1981). This could be related to the fact that the restrictor in their studies was typically only 10 dB below the level at which it would just mask the signal, whereas it was 15 dB below that level in the present study [cf. Figs. 3(b) and (c) in Johnson-Davies and Patterson, 1979]. Further, the signal in Johnson-Davies and

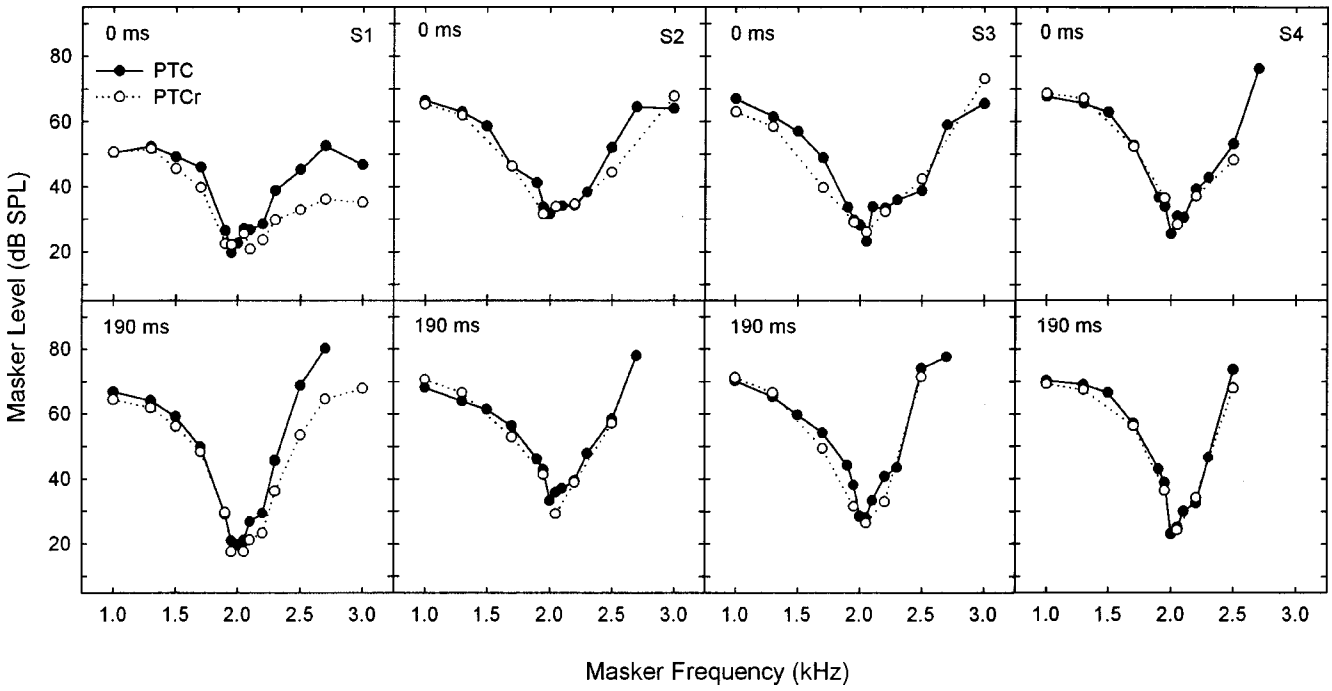


FIG. 2. PTCs from the unrestricted (PTC; replotted from Fig. 1) and restricted (PTCr) paradigms. The top row shows the results for the 0-ms delay, and the bottom row shows the results for the 190-ms delay. The results for individual subjects are shown in separate columns.

Patterson was at a higher SL (almost 20 dB, compared to 10 dB here), and thus it is likely that a greater extent of their signal's excitation pattern would have been usable for off-frequency listening. A comparison across rows in Fig. 2 indicates that the PTCr sharpened with time for each subject, consistent with previous results where a notched noise was used to restrict off-frequency listening (Bacon and Moore, 1986).

The data from the notched-noise experiment are shown in Fig. 3. Masker spectrum level is displayed as a function of the relative deviation of the closer edge of the nearer noise band from f_s. Filled symbols represent the condition where the signal was at the beginning of the masker, and unfilled symbols represent the condition where the signal was at the temporal center of the masker. Circles indicate conditions where the noise was placed symmetrically about f_s, upward-pointing triangles indicate conditions where the higher frequency band was displaced further from f_s than the lower frequency band, and downward-pointing triangles indicate the complementary conditions. All subjects show the same basic trends, though to varying degrees.[1] The effect of signal delay was largest for S1 and S3, smaller but still clearly present for S2, and quite small for S4. Despite the individual differences, all subjects showed little or no difference between the two signal delays for the no-notch condition, and a higher masker level in the 190-ms condition in all the other

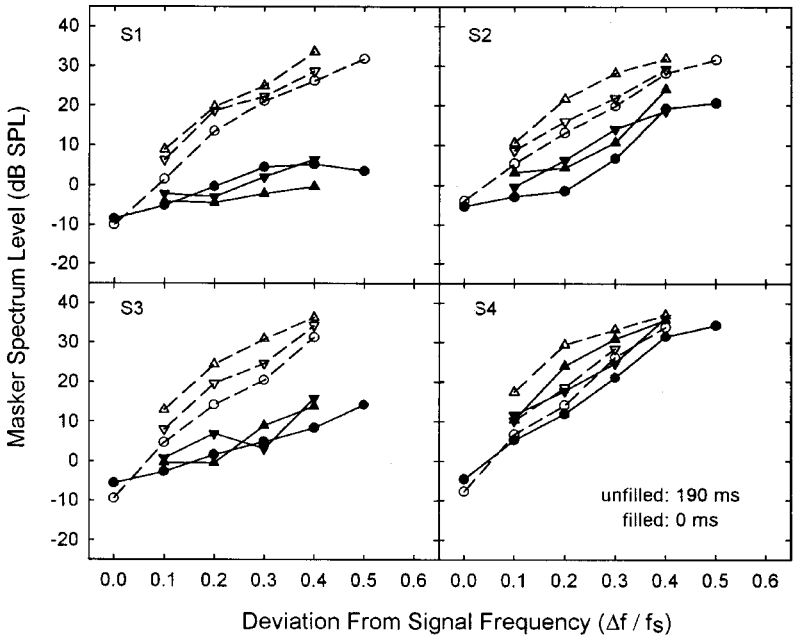


FIG. 3. Level of the notched-noise masker necessary to mask the 10 dB SL signal for a given subject. Filled symbols indicate results for the 0-ms delay; unfilled symbols indicate results for the 190-ms delay. The two noise bands were placed either symmetrically about the 2-kHz signal frequency (circles), or else the upper (triangles) or lower (inverted triangles) band was placed further from the signal. The results are plotted as a function of the relative deviation of the closer edge of the nearer noise band from f_s.

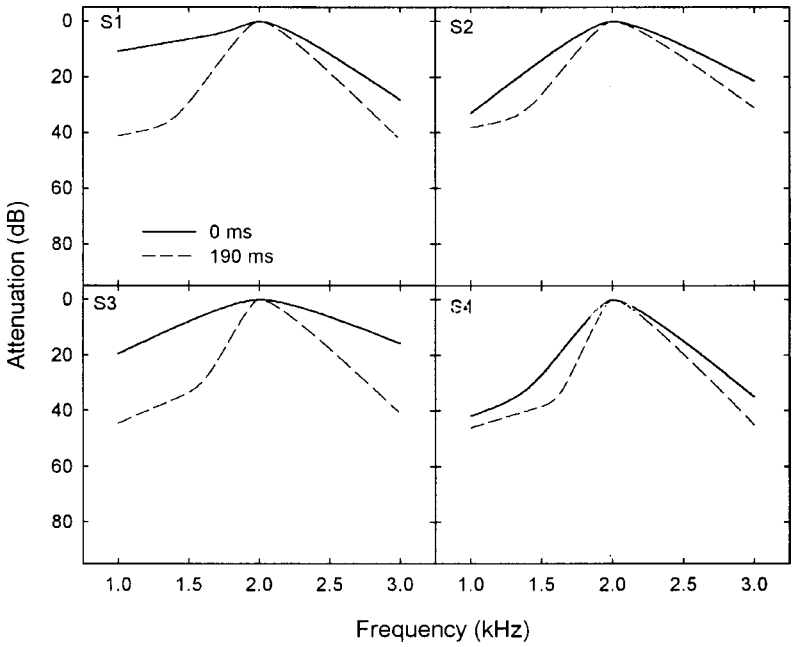


FIG. 4. Auditory filter shapes derived from the data in Fig. 3.

notchwidth conditions (except for a few of the conditions for S4). The difference in masker level between the 0- and 190-ms conditions generally increased with increasing notchwidth. Thus, the functions were steeper with the longer signal delay, suggesting a sharper filter in that condition.

Filter shapes were derived from the notched-noise data for each of the individual subjects using the Polyfit procedure (Rosen *et al.*, 1998) and the roex (*p,w,t*) model given by the following equation (Patterson *et al.*, 1982):

$$W(g)=(1-w)(1+pg)e^{-pg}+w(1+tg)e^{-tg},$$

where p is the parameter that defines the passband of the filter, t defines the tail of the filter, w determines the point where the tail of the filter takes over from the passband, and g is the normalized distance between the center frequency of the filter and the edge of the noise notch. Initially, the parameter values p, w, and t were allowed to differ for the upper (p_u, w_u, t_u) and lower (p_l, w_l, and t_l) sides of the filter. However, the fits to five of the eight data sets were unstable by varying degrees; in other words, several combinations of parameter values yielded nearly equally good fits. The stability was improved considerably by fitting the upper side of the filter with the simple roex (p) shape, thus reducing the number of free parameters used to describe the filter shape from six to four (also see, for example, Rosen *et al.*, 1998). This was achieved with only a small overall change in the goodness of fit (the average rms residual increased from 1.5 to 1.7 dB). With this simpler model, only one fit was unstable (S1, 0-ms condition). Those data were fitted 50 times, and the fit that was obtained most often (37 times) is presented here (a 312-Hz wider filter was obtained the other 13 times; the rms error for that fit was 0.1 dB larger than for the majority fit).

The filter shapes are shown in Fig. 4. The filter for the onset condition (solid line) is clearly broader than that for the centered condition (dashed line). The sharpening with time occurred on both the low- and high-frequency sides. As noted with regard to the data in Fig. 3, there were individual differences in terms of the effect of signal delay; it was large for S1 and S3, and relatively small for S2 and S4. Interestingly, these individual differences are similar to those observed in the PTC paradigm (see Fig. 1). The results in Fig. 4 are inconsistent with those of Moore *et al.* (1987), who did not observe a sharpening with time. They are, however, broadly similar to the results of Wright and Dai (1994).

To compare the tuning revealed in the PTC and notched-noise paradigms, filter shapes were derived from the PTC and PTCr data using the same roex model that was used to derive the filter shapes from the notched-noise data (i.e., using p_l, w_l, t_l, and p_u to define the filter shape). To increase the number of data points on each PTCr, the point for $f_m = f_s = 2$ kHz was taken from the corresponding PTC, under the assumption that off-frequency listening was unlikely to have influenced that point appreciably, if at all. The analyses of the PTC data allowed the filter to shift by an amount $g = 0.1$ (as was done for the notched-noise data); the analyses of the PTCr data did not allow the filter to shift, as it was assumed that the restrictor sufficiently discouraged off-frequency listening. The fits to all the PTC and PTCr data were stable, with the exception of S1 in the 0-ms PTC condition. Those data were fitted 50 times, and the fit that was obtained most often (38 times) is presented here. The other fits were considerably worse, with a rms error that was, on average, 1 dB larger than the majority fit. It is worth noting, however, that even the majority fit was not especially good (see Table I); it is not surprising that the data for this subject (see Figs. 1 and 2) were not fitted well by a filter with the simple roex (p) shape on the high-frequency side.

Figure 5 shows an example PTC (filled circles) and the filter shape (solid line) derived from those data. Both are plotted as attenuation versus frequency, with attenuation increasing along the ordinate. The filter shape is wider than the actual PTC, because the fitting algorithm accounted for off-frequency listening. For this reason, the fits were not as good for the PTC paradigm as for the notched-noise and PTCr

TABLE I. Summary of the parameter values of the best-fitting roex (p,w,t) model to the individual notched-noise (NN), PTC, and PTCr data. The parameters p_l, p_u, t_l, and w_l define the shape of the filter; K refers to the signal-to-masker ratio at the output of the filter required for threshold; rms is a measure of the goodness-of-fit to the data; the ERB was determined by integrating the area under the filter between 0.5 and 4.0 kHz; and $BW_{10\,dB}$ refers to the 10-dB bandwidth of the filter. Results are shown separately for the 0- and 190-ms delays. The mean values are the mean of the individual parameters.

			p_l	p_u	t_l	w_l (dB)	K (dB)	rms (dB)	ERB (Hz)	$BW_{10\,dB}$ (Hz)
0 ms	S1	NN	6.5	20.6	47.8	−144.1	6.0	2.3	576	1360
		PTC	39.0	16.0	0.1	−27.2	2.8	4.7	355	686
		PTCr	33.9	60.0	0.2	−29.4	5.3	1.7	187	359
	S2	NN	19.7	14.4	19.1	−133.9	8.2	2.4	486	944
		PTC	34.5	18.2	6.1	−27.6	−4.7	3.1	338	655
		PTCr	35.1	21.4	6.6	−25.9	−4.5	1.0	306	593
	S3	NN	12.7	10.6	49.9	−78.0	4.8	2.2	671	1305
		PTC	48.4	19.6	11.0	−25.3	1.1	2.9	284	552
		PTCr	28.8	26.5	3.0	−33.0	−0.7	0.9	292	567
	S4	NN	35.1	21.1	10.3	−27.8	5.5	1.5	305	592
		PTC	55.0	30.6	6.1	−34.2	0.5	2.3	202	391
		PTCr	50.4	25.8	4.0	−35.2	−2.7	2.4	234	454
	mean	NN	18.5	16.7	31.8	−96.0	6.2	2.1	510	1050
		PTC	44.2	21.1	5.8	−28.6	−0.1	3.2	295	571
		PTCr	37.0	33.4	3.4	−30.9	−0.6	1.5	255	493
190	S1	NN	37.1	24.7	6.1	−34.0	10.5	1.7	270	524
		PTC	60.0	42.5	8.2	−38.9	9.4	2.5	161	312
		PTCr	55.6	41.1	8.4	−35.7	8.7	2.6	169	328
	S2	NN	34.3	18.9	5.5	−32.0	4.7	0.7	329	638
		PTC	60.0	27.3	8.6	−23.4	−6.4	2.6	217	419
		PTCr	45.2	29.7	8.4	−26.2	−6.5	2.8	223	432
	S3	NN	50.2	24.5	12.4	−27.5	9.9	1.6	247	480
		PTC	60.0	26.9	12.0	−21.8	−4.4	5.8	217	419
		PTCr	47.5	50.4	8.7	−32.9	0.3	2.6	165	320
	S4	NN	58.9	26.3	9.2	−34.6	9.6	1.3	222	431
		PTC	60.0	38.6	4.4	−40.4	1.4	4.9	169	328
		PTCr	60.0	47.3	3.7	−39.6	0.9	3.6	152	294
	mean	NN	45.1	23.6	8.3	−32.0	8.7	1.3	267	518
		PTC	60.0	33.8	8.3	−31.1	0.0	4.0	191	370
		PTCr	52.1	42.1	7.3	−33.6	0.9	2.9	177	344

paradigms (averaged across subjects and delays, the rms residual was equal to 1.7, 3.6, and 2.2 dB for the notched-noise, PTC, and PTCr paradigms, respectively).

As discussed below, the bandwidths of the filters derived from the PTC and PTCr data were similar, and, thus, for clarity, only the former are shown in Fig. 6, along with the filter shapes from the notched-noise data (replotted from Fig. 4). Table I shows the model parameters for the fits to all three sets of data. Included are the four parameters that define the shape of the filter (p_l, w_l, t_l and p_u), K, which defines the processing efficiency following the filter (Patterson *et al.*, 1982), the rms residual, the ERB, and the 10-dB bandwidth. The ERB values are based on an algorithm that determines the area under the derived filter, with integration limits of 0.5 and 4.0 kHz, rather than the formula $[f_s(2/p_l+2/p_u)]$ that is typically used. For the most part, they give the same values, but they can differ considerably when the parameter t for one side of the filter is considerably different from the parameter p on that same side, and when the corresponding parameter w is larger than -20 dB.[2]

Before discussing the filter shapes in Fig. 6, it is worth

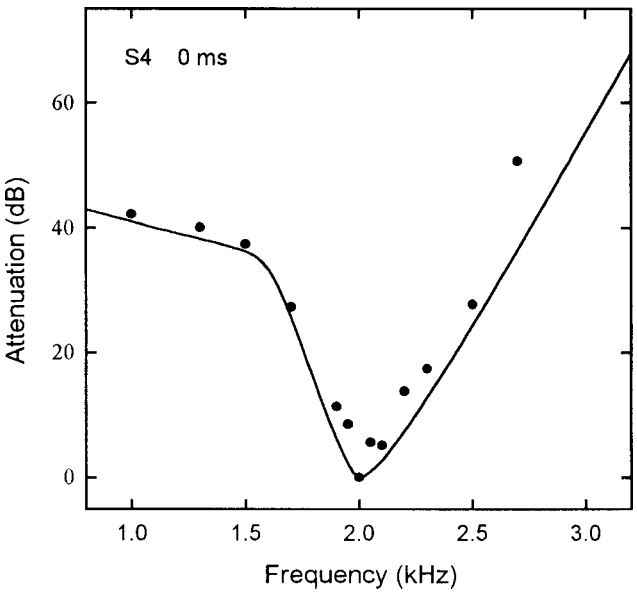


FIG. 5. An example of the roex (p,w,t) fit to a PTC. The symbols represent the PTC from S4 in the 0-ms condition, whereas the solid line represents the derived filter shape.

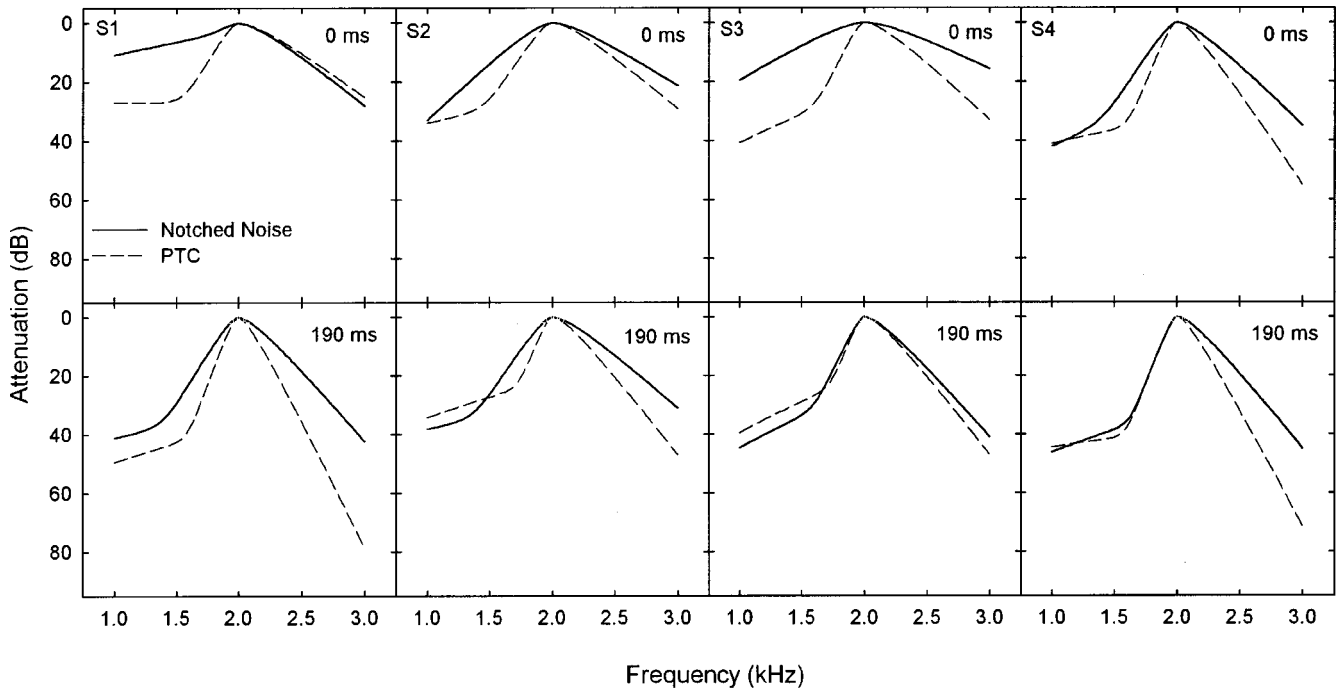


FIG. 6. Filter shapes derived from the PTCs and notched-noise data (replotted from Fig. 4). The top row shows the results for the 0-ms delay, and the bottom row shows the results for the 190-ms delay. The results for individual subjects are shown in separate columns.

noting that K is larger in the notched-noise paradigm than in the PTC or PTCr paradigm. This difference is probably due largely to differences in the variability of the masker itself (see Patterson and Henning, 1977, for a detailed discussion of the relation between K and masker variability). For the two tonal masker paradigms, the mean value of K changed very little as the signal delay increased from 0 to 190 ms (although the values for the individual subjects changed in a way that was inconsistent across subjects). For the notched-noise paradigm, K increased on average by 2.5 dB with increasing signal delay. This seems somewhat paradoxical, in that one might expect K to remain roughly constant or perhaps even decrease with delay (given that the task is generally easier when the signal is at the temporal center of the masker). The increase in K in the notched-noise paradigm, however, presumably reflects the fact that the fitting algorithm must simultaneously account for the following two findings: (1) the filter is narrower at the long delay, and (2) the thresholds are similar in the no-notch condition at the two delays. (See Leek and Summers, 1993, for a detailed discussion of this issue.) Consistent with this is the finding that the increase in K is almost equal to the average decrease in noise power (2.8 dB) expected from the narrower filter in the long-delay condition.

As seen in the top row of Fig. 6, the filter shapes derived from the PTC data (dashed lines) are narrower than those derived from the notched-noise data (solid lines) at the 0-ms delay. In general, the differences between the two filter shapes exist on both the low- and high-frequency sides, although the magnitude of these differences varies across subjects. On average, the ERB and the 10-dB bandwidth values derived from the notched-noise data were about 1.8–2.0 times larger than those derived from the PTC and PTCr data. The bottom row of Fig. 6 shows the filter shapes derived from the two paradigms in the 190-ms condition. The filter shapes from the PTC data are narrower here, too, although the differences between the two filter shapes are not as large as they were at the 0-ms delay. At this longer delay, the differences between the two filter shapes tend to be larger on the high-frequency side. On average, the ERB and the 10-dB bandwidth values derived from the notched-noise data were only about 1.4–1.5 times larger than those derived from the PTC and PTCr data. The average ERB (267 Hz) based on the notched-noise data in the 190-ms condition is similar to that (240 Hz) expected based on other notched-noise data (see Glasberg and Moore, 1990).

The effects of signal delay (0, 190 ms) and masking paradigm (notched noise, PTC, PTCr) were analyzed with a repeated-measures analysis of variance, with the ERB as the dependent variable. There was a significant effect of signal delay [$F(1,3)=21.2$, $p=0.019$], indicating that the filters were narrower in the 190-ms condition. There was also a significant effect of masking paradigm [$F(2,3)=18.6$, $p=0.003$]. *Post-hoc* (Tukey) multiple comparisons analysis ($p<0.05$) indicated that the filters derived from the notched-noise paradigm were significantly wider than those obtained from either the PTC or PTCr paradigm, but that the filter shapes derived from the latter two paradigms did not differ from one another. The interaction between delay and masking paradigm [$F(2,6)=3.3$, $p=0.107$] was not significant, even though the differences between the filters derived from the notched-noise data and those derived from the PTC and PTCr data were reduced at the 190-ms delay. The same significant effects were obtained when the 10-dB bandwidth was the dependent variable.

IV. GENERAL DISCUSSION

A. Changes in tuning with time

The present study was designed to determine whether the tuning revealed in both the PTC and notched-noise paradigms sharpens with time. The results indicate that it does. Although it remains unclear why they sharpen, there are several possible explanations worth considering; most implicate at least a certain amount of central processing. Before doing so, it is important to note that the sharpening observed here is unlikely to represent a spectral artifact resulting from gating the masker. Bacon and Viemeister (1985a) showed that a 10-ms rise/fall time (as used here) is sufficiently long to avoid artificially broad tuning at masker onset due to the splatter of energy from an off-frequency masker to f_s.

The first possible explanation for the sharpening of tuning with time is that it represents some form of adaptation. Short-term adaptation of auditory nerve fibers is often discussed as a potential contributor to temporal effects in simultaneous masking (e.g., Green, 1969; Bacon and Viemeister, 1985a, b; Bacon and Smith, 1991). The response of an auditory nerve fiber declines or adapts during the time in which a relatively long duration stimulus (such as the masker in the present study) is presented, whereas the response to a brief increment in level (like the signal in the present study) does not adapt (Smith and Zwislocki, 1975; Smith, 1979). Consequently, the neural signal-to-masker ratio increases as the signal is delayed from masker onset. However, adaptation in the auditory nerve can only account for about 3 to 5 dB of the temporal effect (Smith and Zwislocki, 1975), and cannot account for the fact that the size of the temporal effect depends upon the frequency relation between masker and signal, given that the amount of adaptation depends only upon the amount by which a fiber is driven by a tone, and not by the frequency of that tone *per se* (Smith, 1979). Thus, peripheral adaptation alone cannot account for a change in tuning with time.

A related possibility, first suggested by McFadden and Wright (1990), is that the sharpening with time is the result of a frequency-specific adaptation followed by an across-frequency comparison. In this scheme, psychophysical masking occurs at a site in the nervous system where comparisons across frequency might occur, and thus at a site central to the auditory periphery [see McFadden and Wright (1990) for a detailed discussion]. Following their argument, the effects of peripheral adaptation will be observed most readily when the masker and signal occupy more or less non-overlapping frequency regions (such as with an off-frequency tonal masker or a notched-noise masker with a relatively wide notch). In that case, the frequency region occupied by the masker will undergo considerably more adaptation than will the region occupied by the signal. Psychophysical tuning will improve with time because the off-frequency masker will need to be increased to compensate for the adaptation.

Another possible explanation is that the change in tuning with time reflects the influence of the efferent system on the outer hair cells (OHCs) in the cochlea. It has been suggested by several investigators that the OHCs may be involved in the temporal effect in simultaneous masking (Kimberley *et al.*, 1989; McFadden and Champlin, 1990; von Klitzing and Kohlrausch, 1994; Bacon and Liu, 2000; Strickland, 2001). Consistent with this possibility are the findings that the temporal effect is reduced or absent in individuals with either a temporary (Champlin and McFadden, 1989; McFadden and Champlin, 1990; Bacon and Hicks, 2000) or permanent (Kimberley *et al.*, 1989; Bacon and Takahashi, 1992; Turner and Doherty, 1997) cochlear hearing loss. Although the OHCs themselves are not thought to change their response properties as a function of time, their effectiveness may be modified by efferent neurons from the medial olivocochlear system. These neurons synapse with OHCs, and the time course of efferent activation is similar to the time course of simultaneous masking (e.g., Warren and Liberman, 1989). Schmidt and Zwicker (1991) and von Klitzing and Kohlrausch (1994) were the first to suggest the possibility that the temporal effect in simultaneous masking might be influenced by the efferent system, and recent psychophysical results from Turner and Doherty (1997), Bacon and Liu (2000), and Zeng *et al.* (2000) are consistent with this possibility (but see Schart *et al.*, 1997). As noted by Strickland (2001), the sharpening of frequency selectivity with time can be understood in terms of a reduction in gain, with time, in the frequency region(s) corresponding to the off-frequency tonal or notched-noise masker, caused by feedback from the efferent system to the OHCs. Thus, as the signal is delayed from masker onset, the masker level will have to be increased to compensate for the decreased gain.

The explanations considered thus far—although plausible—do not necessarily account for why the effect of time was somewhat greater for the notched-noise masker. In particular, although the interaction between delay and masking paradigm was not significant, the filters in the notched-noise paradigm decreased by almost a factor of 2.0, whereas those in the PTC and PTCr paradigms decreased by a factor of only about 1.5. However, it could be argued that, because the efferent effect is stronger when activated by noise than by tones (Warren and Liberman, 1989), the preceding explanation could account for the greater effect of time with the notched-noise masker. This explanation, however, would seemingly predict that, after sufficient time, the tuning should be narrower with the notched-noise masker than with the tonal masker. In other words, it predicts that the greatest differences in tuning with noise and tonal maskers would be in the 190-ms condition, not the 0-ms condition, as observed here. The last two explanations for the change in tuning with time discussed below could explain why the differences in tuning between notched-noise and tonal maskers are greater at masker onset.

The first of these explanations is related to so-called attention filters. In particular, it has been shown that subjects may monitor a wider range of auditory filters when detecting a long duration tone in a gated than in a continuous noise (Dai and Buus, 1991; Wright and Dai, 1994). In other words, their ''attention'' is spread over a wider frequency range when detecting a tone in the presence of a gated masker. Because the psychophysical measure of tuning would be expected to broaden as the number of filters being monitored increased, this may help explain why the auditory filters are

TABLE II. Average standard deviation for each subject in the notched-noise (NN), PTC, and PTCr paradigms. Values are shown separately for the 0- and 190-ms delays. The entries in the rows labeled as ratio are calculated by dividing the standard deviation in the 0-ms condition by the standard deviation in the 190-ms condition; it is an estimate of how the variability changed with signal delay. The last column (S1–S4) represents the average standard deviation across subjects, and represents the variability in thresholds across subjects.

	S1	S2	S3	S4	S1–S4
NN-0	2.5	3.0	3.5	2.3	8.7
NN-190	1.7	2.5	1.9	2.4	3.3
Ratio (NN)	1.5	1.2	1.9	0.9	2.6
PTC-0	3.7	2.4	3.0	2.0	6.1
PTC-190	2.3	1.7	2.3	2.0	4.9
Ratio (PTC)	1.7	1.4	1.3	1.0	1.2
PTCr-0	3.3	1.6	2.5	2.4	5.8
PTCr-190	2.9	1.0	1.5	1.7	4.6
Ratio (PTCr)	1.1	1.6	1.7	1.4	1.3

wider at masker onset. Although attention filters have not been measured using maskers and signals with the same spectral and temporal characteristics as those employed in the present study, it is possible that the subjects tested here monitored a wider range of filters when detecting the 20-ms signal at the beginning of the masker than they did when detecting the signal later in time, when the effects of gating on attention probably would have subsided, at least somewhat. Furthermore, the relatively broad noise masker may have drawn the subjects' attention over a wider frequency range than did the tonal masker, which could account for why tuning at masker onset was especially poor in the notched-noise paradigm.

Finally, another possible explanation for the broader tuning at masker onset is that subjects may have difficulty estimating the spectrum of the masker at its onset, which in turn could contribute to particularly poor frequency selectivity in the 0-ms condition. Because noise is inherently variable, and because the spectrum of the notched-noise masker is more complex than that of the tonal masker, it may be more difficult to estimate the spectrum of the noise masker than that of the tonal masker, which may be why tuning was especially poor in the 0-ms condition with the notched-noise masker. For both tonal and noise maskers, subjects would likely improve their estimate over time as a result of more ''looks'' at the masker. The presumed difficulty in estimating the masker's spectrum might be manifest as a particularly variable threshold in the 0-ms condition. A measure of this variability is shown in Table II, where the average standard deviation for each subject at each delay was calculated for the notched-noise, PTC, and PTCr paradigms (the last column, labeled S1–S4, will be described later). This was done by averaging the standard deviations for each of the conditions within a paradigm. (Recall that the standard deviation for each condition is based on averaging three or more threshold estimates for that condition.) The average standard deviation was almost always larger in the 0-ms condition than in the 190-ms condition. If the broader tuning at masker onset is related to the larger variability in that condition, then one might expect that individual differences in the sharpening with time would be mirrored by individual differences in the decreases in variability with time. Consistent with this possibility is the fact that, for the notched-noise and PTC paradigms, the rank ordering of subjects in terms of the percentage change in tuning was the same as that in terms of the percentage change in variability, as defined by the ratio of the average standard deviations in the two delay conditions. For the PTCr paradigm, this was almost the case: the first and last in the rank were the same for the change in tuning and the change in variability, but the second and third in rank were reversed. In addition, if variability at masker onset is related to broad tuning, then one might expect the variability in the onset condition to be greatest in the notched-noise paradigm. This was the case for S2 and S3. Overall, then, there is reasonable support for a relation between variability (possibly reflecting a difficulty in estimating the spectrum of the masker) and broad tuning at masker onset.

In addition to the variability being greater when the signal was at the beginning of the masker for all masker paradigms (the ratios in Table II are almost all greater than 1.0), thresholds differed more across subjects in the 0-ms condition than in the 190-ms condition (see Figs. 1–3). This is revealed in the last column of Table II. This standard deviation represents the intersubject variability averaged across all conditions within a paradigm, where the standard deviation for a given condition was obtained by averaging all the estimates (3 or more per subject$\times$4 subjects) for that condition. These standard deviations are considerably larger in the 0-ms condition (especially in the notched-noise paradigm), indicating that thresholds differed more across subjects in that condition than in the 190-ms condition. The greater intersubject (and intrasubject) variability for a signal at masker onset suggests that this condition is inherently more difficult than the one where the signal is at the temporal center of the masker. This, in turn, is consistent with the possibility that central processing may have a considerable influence on the measures of tuning at onset.

B. Comparisons of tuning among paradigms in the 190-ms condition

The PTC and the filter shape derived from notched-noise data are the two most common psychophysical measures of auditory tuning. The present study is apparently the first to compare PTCs and filter shapes from notched-noise data in the same subjects. Further, it is the first to derive filter shapes from PTC data, thus allowing a more direct comparison of the tuning revealed by the two paradigms. As discussed above, the filters from the notched-noise paradigm were significantly wider than those from the PTC and PTCr paradigms, and although there was not a statistical interaction between delay and masking paradigm, the differences in filter bandwidth were somewhat larger in the 0-ms condition (the filter derived from notched-noise data was about 2 times wider than those derived from the tonal data at the 0-ms delay, and 1.5 times wider at the 190-ms delay). The previous section discussed possible explanations for this, with most of them focusing on processing at a central level of the auditory system. This section focuses only on comparisons in

the 190-ms condition. This is a steady-state condition, which is the type of condition most commonly used to evaluate measures of frequency selectivity.

Before attempting to explain the differences in filter shape revealed in the bottom row of Fig. 6, it is worth noting that they are unlikely due to a number of factors that *could* influence the comparisons between the tuning revealed by the PTC and notched-noise paradigms. These include combination tones, beats, and off-frequency listening. The detection of a combination product, which is more likely to occur in the PTC paradigm, can result in an artificially sharp PTC, especially on the low-frequency side (Nelson and Fortune, 1991). However, a low-pass noise was used in the present study to mask the CDT; furthermore, at such low levels, the detection of combination tones is unlikely (Smoorenburg, 1972). The detection of beats can also influence the shape of the PTC, particularly around its tip, but is unlikely to influence the filter derived from notched-noise data. The short duration signal used here, however, probably minimized the detection of beats in the present study. Finally, as discussed previously, off-frequency listening is more likely to affect the PTC, making it too sharp (Johnson-Davies and Patterson, 1979). Because off-frequency listening in the PTC paradigm was taken into account, either with a restrictor or by fitting the roex model to the data (and allowing the filter to shift), it is unlikely that it had a significant impact on the comparisons made here.

Across the four subjects, the largest differences between the filters derived from notched-noise and PTC data were on the high-frequency side (bottom row of Fig. 6). One possible explanation for this difference is suppression. It has been suggested that measures of tuning are broader in simultaneous than in nonsimultaneous masking because of suppression by off-frequency maskers (Houtgast, 1973; Moore, 1978). These differences in tuning are particularly evident on the high-frequency side. Because the strength of suppression may increase with the number of components in the masker (Lee and Bacon, 1998), it is possible that suppression had a greater effect in the notched-noise paradigm, resulting in a broader filter on the high-frequency side. Consistent with the present results—where the filter from the noise masker is broader on the high-frequency side—is the fact that, at low levels, such as those used here, suppression is more effective from the high-frequency side (e.g., Houtgast, 1973). Because suppression presumably follows filtering in the sequence of auditory events, this explanation implicates processing that is at least somewhat central to auditory filtering (even if it exists in the auditory periphery). The suppression explanation could be evaluated by comparing filters derived from these same paradigms under conditions of nonsimultaneous masking.

V. SUMMARY

Auditory filter shapes were derived from PTC, PTCr, and notched-noise data using the roex (p,w,t) model. The 20-ms signal was fixed in level at 10 dB SL and presented at the onset or temporal center of the 400-ms masker. The masker was varied adaptively to mask the 2-kHz signal. The main findings can be summarized as follows:

(1) The filters from all three paradigms sharpened with time.
(2) The change in tuning with time was somewhat greater in the notched-noise paradigm than it was in the PTC and PTCr paradigms.
(3) The filter from the notched-noise paradigm was significantly wider than that from the PTC and PTCr paradigms, especially at masker onset, where the differences may reflect the influences of central processing. In general, at masker onset, the filter derived from the notched-noise data was wider on both the low- and high-frequency sides.
(4) When the signal was at the temporal center of the masker, the filter derived from the notched-noise paradigm was primarily wider on the high-frequency side. This difference in filter bandwidth might reflect the influence of suppression, which can be greater with noise than with tonal maskers.
(5) Finally, although the notched-noise results from Moore *et al.* (1987) are in disagreement with those presented here (and elsewhere), one of the conclusions from the present study is in agreement with their conclusion that the temporal effect with PTCs is due to some aspect other than auditory filtering *per se*. Indeed, the temporal effect with both tonal and notched-noise maskers probably reflects the influence of central processing, which may differ somewhat for the two types of masker.

ACKNOWLEDGMENTS

This research was supported by NIDCD Grant No. DC 01376. It is based on a Master's thesis submitted by the second author to the Graduate College at Arizona State University. We would like to thank the reviewers for their comments on an earlier version of this manuscript. Reviewer Roy Patterson in particular provided numerous helpful suggestions, including the one to fit the PTCs with the roex model and the one to analyze variability in the context of understanding the temporal effect. Finally, we also would like to thank Richard Baker for his generous help with PolyFit, particularly in fitting the PTC and PTCr data.

[1]One unusual trend can be seen in the data for S1 in the 0-ms condition. In some instances, for a given deviation, the masker level was lower in an asymmetric condition (especially when the higher frequency band was further away) than in the symmetric condition. In other words, a larger notch resulted in poorer signal detectability, which cannot be accounted for with the power spectrum model of masking. Given the low SL and wide frequency separations at which this anomalous result occurred, it is extremely unlikely to have resulted from the detection of combination products (Smoorenburg, 1972). It may instead reflect the generally variable thresholds coupled with the extremely shallow function relating masker level to notchwidth. For this subject and delay, masker level changed by less than 14 dB with changes in notchwidth. Nevertheless, the roex model fitted these data about as well as any of the other data in the 0-ms condition (see rms values in Table I).

[2]For particularly wide filters, where the tail takes over from the passband at attenuation values less than 20 dB, the ERB value based on the formula $f_s(2/p_l+2/p_u)$ will not be accurate (except when $t=p$). This could occur, for example, at high levels or in individuals with considerable hearing loss.

ANSI (**1996**). ANSI S3.6-1996, ''Specifications for audiometers'' (American National Standards Institute, New York).
Bacon, S. P., and Hicks, M. L. (**2000**). ''The influence of aspirin on temporal effects in simultaneous masking with noise and tonal maskers,'' J. Acoust. Soc. Am. **107**, 2914.
Bacon, S. P., and Liu, L. (**2000**). ''Effects of ipsilateral and contralateral precursors on overshoot,'' J. Acoust. Soc. Am. **108**, 1811–1818.
Bacon, S. P., and Moore, B. C. J. (**1986**). ''Temporal effects in masking and their influence on psychophysical tuning curves,'' J. Acoust. Soc. Am. **80**, 1638–1645.
Bacon, S. P., and Smith, M. A. (**1991**). ''Spectral, intensive, and temporal factors influencing overshoot,'' Q. J. Exp. Psychol. A **43**, 373–399.
Bacon, S. P., and Takahashi, G. A. (**1992**). ''Overshoot in normal-hearing and hearing-impaired subjects,'' J. Acoust. Soc. Am. **91**, 2865–2871.
Bacon, S. P., and Viemeister, N. F. (**1985a**). ''Simultaneous masking by gated and continuous sinusoidal maskers,'' J. Acoust. Soc. Am. **78**, 1220–1230.
Bacon, S. P., and Viemeister, N. F. (**1985b**). ''The temporal course of simultaneous tone-on-tone masking,'' J. Acoust. Soc. Am. **78**, 1231–1235.
Carlyon, R. P. (**1989**). ''Changes in the masked thresholds of brief tones produced by prior bursts of noise,'' Hear. Res. **41**, 223–235.
Champlin, C. A., and McFadden, D. (**1989**). ''Reductions in overshoot following intense sound exposures,'' J. Acoust. Soc. Am. **85**, 2005–2011.
Dai, H., and Buus, S. (**1991**). ''Effect of gating the masker on frequency-selective listening,'' J. Acoust. Soc. Am. **89**, 1816–1818.
Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.
Glasberg, B. R., Moore, B. C. J., and Nimmo-Smith, I. (**1984a**). ''Comparison of auditory filter shapes derived with three different maskers,'' J. Acoust. Soc. Am. **75**, 536–544.
Glasberg, B. R., Moore, B. C. J., Patterson, R. D., and Nimmo-Smith, I. (**1984b**). ''Dynamic range and asymmetry of the auditory filter,'' J. Acoust. Soc. Am. **76**, 419–427.
Green, D. M. (**1969**). ''Masking with continuous and pulsed sinusoids,'' J. Acoust. Soc. Am. **46**, 939–946.
Greenwood, D. D. (**1971**). ''Aural combination tones and auditory masking,'' J. Acoust. Soc. Am. **50**, 502–543.
Hawkins, J. E., and Stevens, S. S. (**1950**). ''The masking of pure tones and of speech by white noise,'' J. Acoust. Soc. Am. **22**, 6–13.
Hicks, M. L., and Bacon, S. P. (**1992**). ''Factors influencing temporal effects with notched-noise maskers,'' Hear. Res. **64**, 123–132.
Houtgast, T. (**1973**). ''Psychophysical experiments on 'tuning curves' and 'two-tone inhibition,' '' Acustica **29**, 168–179.
Johnson-Davies, D., and Patterson, R. D. (**1979**). ''Psychophysical tuning curves: Restricting the listening band to the signal region,'' J. Acoust. Soc. Am. **65**, 765–770.
Kimberley, B. P., Nelson, D. A., and Bacon, S. P. (**1989**). ''Temporal overshoot in simultaneous-masked psychophysical tuning curves from normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **85**, 1660–1665.
Lee, J., and Bacon, S. P. (**1998**). ''Psychophysical suppression as a function of signal frequency: Noise and tonal maskers,'' J. Acoust. Soc. Am. **104**, 1013–1022.
Leek, M. R., and Summers, V. (**1993**). ''Auditory filter shapes of normal-hearing and hearing-impaired listeners in continuous background noise,'' J. Acoust. Soc. Am. **94**, 3127–3137.
Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Lutfi, R. A., and Patterson, R. D. (**1984**). ''On the growth of masking asymmetry with stimulus intensity,'' J. Acoust. Soc. Am. **76**, 739–745.
McFadden, D., and Champlin, C. A. (**1990**). ''Reductions in overshoot during aspirin use,'' J. Acoust. Soc. Am. **87**, 2634–2642.
McFadden, D., and Wright, B. A. (**1990**). ''Temporal decline of masking and comodulation detection differences,'' J. Acoust. Soc. Am. **88**, 711–724.
Moore, B. C. J. (**1978**). ''Psychophysical tuning curves measured in simultaneous and forward masking,'' J. Acoust. Soc. Am. **63**, 524–532.
Moore, B. C. J., Poon, P. W. F., Bacon, S. P., and Glasberg, B. R., (**1987**). ''The temporal course of masking and the auditory filter shape,'' J. Acoust. Soc. Am. **81**, 1873–1880.
Nelson, D. A., and Fortune, T. W. (**1991**). ''High-level psychophysical tuning curves: Simultaneous masking by pure tones and 100-Hz-wide noise bands,'' J. Speech Hear. Res. **34**, 360–373.
O'Loughlin, B. J., and Moore, B. C. J. (**1981**). ''Off-frequency listening: Effects on psychoacoustical tuning curves obtained in simultaneous and forward masking,'' J. Acoust. Soc. Am. **69**, 1119–1125.
Patterson, R. D., and Henning, G. B. (**1977**). ''Stimulus variability and auditory filter shapes,'' J. Acoust. Soc. Am. **62**, 649–664.
Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). ''The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold,'' J. Acoust. Soc. Am. **72**, 1788–1803.
Rosen, S., and Baker, R. J. (**1994**). ''Characterising auditory filter nonlinearity,'' Hear. Res. **73**, 231–243.
Rosen, S., Baker, R. J., and Darling, A. (**1998**). ''Auditory filter nonlinearity at 2 kHz in normal hearing listeners,'' J. Acoust. Soc. Am. **103**, 2539–2550.
Scharf, B., Magnan, J., and Chays, A. (**1997**). ''On the role of the olivocochlear bundle in hearing: 16 case studies,'' Hear. Res. **103**, 101–122.
Schmidt, S., and Zwicker, E. (**1991**). ''The effect of masker spectral asymmetry on overshoot in simultaneous masking,'' J. Acoust. Soc. Am. **89**, 1324–1330.
Smith, R. L. (**1979**). ''Adaptation, saturation, and physiological masking in single auditory-nerve fibers,'' J. Acoust. Soc. Am. **65**, 166–178.
Smith, R. L., and Zwislocki, J. J. (**1975**). ''Short-term adaptation and incremental responses of single auditory-nerve fibers,'' Biol. Cybern. **17**, 169–182.
Smoorenburg, G. F. (**1972**). ''Audibility region of combination tones,'' J. Acoust. Soc. Am. **52**, 603–614.
Strickland, E. A. (**2001**). ''The relationship between frequency selectivity and overshoot,'' J. Acoust. Soc. Am. **109**, 2062–2073.
Turner, C. W., and Doherty, K. A. (**1997**). ''Temporal masking and the 'active process' in normal and hearing-impaired listeners,'' in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Hillsdale, NJ), pp. 387–396.
von Klitzing, R., and Kohlrausch, A. (**1994**). ''Effect of masker level on overshoot in running- and frozen-noise maskers,'' J. Acoust. Soc. Am. **95**, 2192–2201.
Warren III, E. H., and Liberman, M. C. (**1989**). ''Effects of contralateral sound on auditory-nerve responses. I. Contributions of cochlear efferents,'' Hear. Res. **37**, 89–104.
Wright, B. A., and Dai, H. (**1994**). ''Detection of unexpected tones in gated and continuous maskers,'' J. Acoust. Soc. Am. **95**, 939–948.
Zeng, F.-G., Martino, K. M., Linthicum, F. H., and Soli, S. D. (**2000**). ''Auditory perception in vestibular neurectomy subjects,'' Hear. Res. **142**, 102–112.
Zwicker, E. (**1979**). ''Different behavior of quadratic and cubic difference tones,'' Hear. Res. **1**, 283–292.

The attention filter for tones in noise has the same shape and effective bandwidth in the elderly as it has in young listeners

James R. Ison,[a] Tracy M. Virag, and Paul D. Allen
Department of Brain and Cognitive Sciences, Meliora Hall, University of Rochester, Rochester, New York 14627

Geoffrey R. Hammond
Department of Psychology, Psychology Building, University of Western Australia, Nedlands, Western Australia 6907

(Received 19 September 2001; revised 11 April 2002; accepted 15 April 2002)

Listeners asked to detect tones masked by noise hear frequent signals but miss infrequent probes, suggesting that they attend to spectral regions where they expect the signals to occur. The narrow detection pattern centered on the frequent target approximates that obtained in notched noise, indicating that attention is focused on the auditory filter. We measured attention bands in young and elderly listeners ($n=5$, 4; 20–25 and 62–82 years of age) for targets (800 or 1200 Hz) and infrequent probe signals (target ± 25–100 Hz) masked in wideband noise. We anticipated that their width would increase with age, as has been reported for auditory filters. A yes–no single-interval procedure provided detection probabilities and detection response speeds. Both measures showed near-linear declines with decreasing signal level, and graded decay functions as probe frequency deviated from the target frequency. Probes deviating from the target by 25 to 50 Hz were equivalent to a 2-dB reduction in signal level for both measures. The equivalent rectangular bandwidth (ERB) for detection approximated 11% of the signal frequency for each age group. Confidence intervals (95%) showed that the elderly ERB could be at most only about 20% larger than that of younger listeners. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1483321]

PACS numbers: 43.66.Lj, 43.66.Dc, 43.66.Sr [MRL]

I. INTRODUCTION

Stimulus detection is invariably difficult in the presence of masking noise, but this adverse effect is especially noticed by elderly listeners (Working Group on Speech Understanding and Aging, 1988). Masking is reduced in its severity when the spectral features of the signal are known to the listener, who can then selectively listen for these particular features while ignoring spectral regions in the masker where the signal is less likely to occur. The present experiment examined the hypothesis that the specificity of this beneficial listening strategy is diminished in the elderly. Our procedure depends on a classic observation in the psychophysical laboratory, namely, that auditory sensitivity depends in part on the recent history of stimulus presentation, so that repeated and/or recent exposure to a particular signal frequency leads to improved detection for a narrow band of tones centered on that signal (Swets *et al.*, 1959; Tanner *et al.*, 1956, cited in Swets *et al.*, 1959). Greenberg and Larkin (1968) measured the shape and the spectral extent of the selective listening band by examining the effect on probe detectability of varying the relative occurrence of off-frequency probes as well as the degree to which they deviated from the tonal frequency of the target signal. The consistent outcome of the several procedures they explored through their series of experiments was that the more common target signals were detected at a high probability while the less common probes were detected less often: indeed, the most deviant off-frequency probes were detected only at a chance level even though all of the tones were presented at equivalent intensity levels. The response patterns that map stimulus detection onto spectral deviation of the probes from the target signal have been called ''attention bands'' (Scharf *et al.*, 1987, p. 222) or ''listening bands'' (Schlauch and Hafter, 1991, p. 1332). These bands appear to reflect the development of listening strategies that are selected by the subjects, not necessarily intentionally, in order to maximize the detection of the expected target signals.

Swets and Kristofferson (1970) acknowledged that these band-pass filters could result from a listener's choosing not to report the presence of unfamiliar probes rather than a failure to hear them, but Scharf and his colleagues have more recently provided two persuasive lines of evidence supporting the hypothesis that the listeners do not hear the off-frequency probe tones. Scharf *et al.* (1987) found that informing the listeners that different tones would be presented across trials did not increase the detectability of unfamiliar probes, and also that probes that ''sounded the same'' as the target signals because of shared harmonic and octave relationships were not reported, while complex probes that ''sounded different'' but contained the same critical signal frequencies were detected. This latter effect further suggested to Scharf *et al.* (1987) that the attentional mechanism is focused on the sensory receptors in the cochlea, certainly below the higher auditory centers at which it has been sug-

[a]Electronic mail: ison@bcs.rochester.edu

 0001-4966/2002/112(1)/238/9/$19.00

gested that harmonic and octave relationships are appreciated (Houtsma, 1995).

That the neural target of frequency selectivity is the sensory receptor is most strongly supported by two subsequent reports showing that sharply tuned attention bands are absent in patients who lack efferent control over the sensitivity of the hair cell along the basilar membrane (Scharf *et al.*, 1994, 1997). Efferent control of receptor sensitivity is provided by the olivocochlear bundle (OCB), which runs out from the superior olivary nuclei of the auditory brainstem through the vestibular branch of the statoacoustic nerve to terminate directly on the outer hair cells or on the afferent processes that innervate the inner hair cells (see, e.g., Warr *et al.*, 1986; Wiederhold, 1986). A surgical treatment for unilateral Meniere's disease is to cut the vestibular branch on the affected side, and thus its incidental consequence is to eliminate efferent control over receptor sensitivity normally provided by the OCB (as described in Scharf *et al.*, 1994, p. 14). While the effect of this surgery is minimal for most measures of hearing, an outcome that is critical in the present theoretical context is that these patients show a substantial increase in their detection of off-frequency probes in the probe-signal experiment. This effect is shown in comparisons of preoperative versus postoperative performance (see Fig. 5, Scharf *et al.*, 1994), or postoperative performance alone, in comparisons of listening to masked tones in the intact versus the impaired ear (see Fig. 5, Scharf *et al.*, 1997). In either case, when masked tones are presented to an ear that lacks OCB input, then all but the most deviant off-frequency probes are detected with the same high probability found for the target signal, while probe presentations in the ear with an intact OCB are not reported as being heard. These data are consistent with the idea that the neural target of the selective mechanism is the peripheral receptor. They also support the argument that the frequency selectivity obtained in these experiments results because the OCB normally attenuates receptor sensitivity at the probe frequencies rather than its being the product of a cautious reporting strategy: a cautious listener would be expected to censor any reports of hearing deviant tones regardless of the ear in which those stimuli had been presented.

The rationale for the present comparison of the attention band in elderly versus young listeners is based on these reports plus data described below showing that other aspects of OCB function change in elderly listeners, these suggesting that age-related failures in selective attention may contribute to their difficulty in understanding speech in noisy environments. Collet *et al.* (1990) showed in human listeners that transient otoacoustic emissions evoked by tone pips in one ear could be suppressed by presenting noise to the contralateral ear, and Castor *et al.* (1994) and Kim *et al.* (2002) showed that this suppressive effect is diminished in elderly listeners. It is understood that these emissions are the product of outer hair cell activity in response to the tone (Kemp, 1978), and that suppression results because the contralateral noise engages the olivocochlear system: the contralateral branch of the OCB crosses the midline and terminates on the outer hair cells, and then directly inhibits outer hair cell activity in the ear from which the emissions are recorded (Liberman, 1989). Thus the data reported by Castor *et al.* (1994) and Kim *et al.* (2002) suggest that there is an age-related loss in OCB function that might also be observed in the selective listening experiment. Adding further to the interest in these effects are data showing a relationship between OCB activity and phoneme recognition in noise (Giraud *et al.*, 1997). This work, together with the experiments of Castor *et al.* (1994) and Kim *et al.* (2002), showing age-related deficits in one aspect of OCB functioning, and the reports of Scharf and his colleagues (1994, 1997), showing the importance of the OCB for selective attention to masked stimuli, jointly suggest the hypothesis that a loss of selective attention may contribute to speech recognition deficits in the elderly presbycusic listener. This hypothesis gains some further support in demonstrations that the passband of auditory filters measured in notched noise increases with age (Patterson *et al.*, 1982), as at least in young normal listeners attention bands approximate the width of the auditory filter (Greenberg and Larkin, 1968, and others). More direct evidence for this hypothesis would be provided by a demonstration that elderly listeners do show relatively broad attention bands as assessed by the probe-signal method.

Our variant of this paradigm used a modified method of constant stimuli with a yes–no single interval procedure. The target signals were all at one fixed frequency, but were presented at a range of levels at a set time delay after noise onset. On some relatively rare occasions probe stimuli having different spectral frequencies were substituted for the target signal, these all presented at one fixed intermediate level. The reason for using this ''yes–no'' single interval experimental design was that it allowed us to measure detection response times as well as detection probability. It has been known for a very long time that simple reaction times are faster when the observer expects a particular stimulus to occur at a particular time (James, 1890, Vol. 1, p. 425; Woodworth and Schlosberg, 1955, p. 28). Simple reaction times for detecting expected target signals versus infrequent probes have apparently not been examined in the signal-probe research design, save for one negative finding briefly described by Scharf (1998, p. 98) in an experiment that used target and probe signal levels a little higher than is typical for these experiments. Some studies of auditory attention have shown that being able to correctly anticipate the spatial location of a target tone increases the speed with which one tonal pitch can be discriminated from another in a disjunctive reaction time test (Mondor and Zatorre, 1995; Scharf, 1998; Spence and Driver, 1994, 1996), but this information seems not to affect simple detection reaction times, so that, e.g., tones appearing at expected and unexpected locations are responded to with equal rapidity (Buchtel and Butter, 1988; Scharf, 1998; Spence and Driver, 1994, 1996). Scharf (1998) described these effects, and suggested that the reason that the manipulations of spatial attention that substantially affect detection speeds for visual experiments have little effect in auditory experiments, first noted by Posner (1978), is based on a fundamental difference in the initial encoding of photic and acoustic events. The visual stimulus is always fixed at a particular site on the retina (in the absence of eye and head movements, of course), while the representation of the audi-

tory stimulus is first diffusely distributed across the basilar membrane, and then contracts with an increase in its duration. One necessary consequence of this shift in the sensory-neural representation of tonal stimuli with their continued presentation is that the pitch of brief tones must be less salient than that of sustained tones: thus, for example, von Bekesy (1960/1929, p. 221) reported that two-tone discrimination progressively diminished at durations below 100 ms, and Moore (1973) showed that the frequency difference limen progressively increased as tone duration fell from 200 to 6.25 ms. It seemed thus possible to Scharf (1998) that the onset of a tone could be detected and responded to in a simple reaction time task before its tonal quality could be registered at the binaural neural mechanisms responsible for locating objects in space on which attention has been focused. This is a plausible explanation of the failure of attention to spatial location to affect simple reaction times, perhaps especially given the presentation of a high-level suprathreshold stimuli typical of the spatial attention paradigm. However, we suggest that a different outcome can be expected when attention in directed to *spectral* location in the probe-signal paradigm if the listening band is shaped at the level of the basilar membrane by centrally induced OCB activation, as the reports of Scharf and his colleagues described above indicate. In this paradigm the graded attenuation of hair cell sensitivity along the basilar membrane is assumed to be in place in preparation for the presentation of the typically near-threshold tonal signal in the noise background, and thus the fact that its initial representation may be relatively diffuse would not increase its detectability. Thus we anticipate that detection probability and detection speed should decline with increasing distance between the probes and the target signal, both showing the same degree of effective attenuation when mapped onto the detection rates and speeds of targets presented at lower levels.

To anticipate the outcome of this study, the hypothesis that stimulus selection should be equally effective in diminishing both probability and speed measures of detection was confirmed, but the hypothesis that the attentional mechanism should be less effective in the elderly listener was not: the shapes and bandwidths of the auditory attention filters were virtually identical in the two age groups.

II. METHOD

A. Subjects

Five young listeners (3F, 2M, ages 20 to 25 years) and four elderly listeners (3M, 1F, ages 62, 65, 67, and 82 years) took part in the experiment. The younger listeners were advanced undergraduate and graduate students at the University of Rochester who were recruited from an advertising poster, and the elderly listeners were working academics or retired academics still active in the community who were recruited by personal contact. All had their hearing tested before the beginning of the experiment. The young had normal hearing within ± 10 dB HL (ANSI, 1989) at all frequencies, and none complained of any hearing problem. The elderly had normal hearing for their age (see Sec. III).

B. Apparatus and stimuli

Hearing thresholds were assessed using a Bausch & Lomb Audiorater III with TDH-39P headphones, at eight frequencies: 125, 250, 500, 1000, 1500, 2000, 4000, and 8000 Hz. The stimuli in the attention paradigm included wideband noise and tone bursts, generated by a digital signal-processing board [Tucker Davis Technologies (TDT) Array Processor 2 (AP2)]. The noise bursts were comprised of 3-s segments (including 5 ms linear rise–fall times) that were generated randomly by the AP 2 prior to the beginning of each observation period, transduced by a 16-bit D/A converter (TDT DA 1) and attenuated by a programmable attenuator (TDT PA 4) to an overall 60 dBA SPL. The noise bursts were routed through a signal mixer (TDT SM 3) to a headphone buffer (TDT HB 5) and presented binaurally over Beyer earphones (Beyer DT 48, having a relatively flat frequency response through 12 kHz) that were held in a circumaural cushion. A tone was presented a constant 2 s after noise onset on 65 of the 70 trials that comprised each observation block (the remaining 5 trials were no-stimulus catch trials). The tones were 200 ms in duration with linear 5 ms rise/fall times, and ranged in level from 47 to 55 dB dBA SPL in 2-dB steps. The target signal tone with a high probability of occurrence was either 800 Hz (for $n=3{:}1$ old, 2 young) or 1200 Hz ($n=6{:}3$ old, 3 young), these being assigned at random across listeners in order to minimize the effects of any idiosyncratic spectral perturbations in the response characteristics of the earphones at the target signal and probe frequencies. It should be noted, however, that any peculiar uneven spot in the sensitivity of the earphone in the spectral regions of interest in this experiment would have matched effects on the stimulus tones and on the surrounding masking noise, and this should thus preserve the local signal-to-noise ratio across listeners. There were eight infrequent probe tones, spaced at four 25-Hz intervals for 100 Hz on either side of the frequent signal, and always presented at 49 dBA SPL, the same as one of the target signal levels. The tones were also generated by the AP 2 card, transduced by a second D/A converter, attenuated by a second PA 4, added to the noise through the signal mixer and the headphone buffer, and finally delivered binaurally to the earphones. Responses were measured by the listeners' pressing one of two raised response buttons on a small aluminum box ($18\times 13\times 5\ \mathrm{cm}^3$), and were recorded to the nearest ms, the time interval beginning with the onset of the tone. The response buttons, made of translucent red plastic that could be illuminated, were placed 3 cm in from the sides on the longer axis, so that they could be easily depressed with the right or left thumb. Stimulus presentation and response measurement were under computer control.

C. Procedure

All of the procedures were approved by the Research Subjects Review Board of the University of Rochester. The participants were first informed about the general conditions of the experiment, after which they read and signed the consent form. Their hearing was then measured and the experimental sessions began. The listeners sat comfortably in a

double-walled sound booth (Industrial Acoustics Corporation, approx. $2\times2\times2$ m^3), holding the response box. They were informed that different tones would be presented in the noise always at one particular interval, but that some trials would be blank. If they heard a tone they were to press the right button as quickly as possible, but if the noise ended and they had not heard the tone, then they were to press the left button to indicate that they had not heard a tone on that trial. If they pressed the ''Yes'' button on a blank trial, then that button would be briefly illuminated. They worked for 1 h each day for 2 to 5 days, this depending on their availability, and for each hour they were paid $7.50. On each day they received six blocks of trials separated by a 2–5-min break. Each block began with a series of ''warm-up'' trials in which only the target signal tone was presented at 49, 51, 53, and 55 dBA, each three times in random order within blocks of trials. This introduction led seamlessly into the experimental series of trials which contained a total of ten presentations each of the target signal tone (i.e., 800 or 1200 Hz for different listeners) at 47, 51, 53, and 55 dBA; five presentations each of the target signal and a selection of four of the eight possible off-frequency probes (i.e., spanning the range of 700 to 900 Hz, or 1100 to 1300 Hz) all at 49 dBA; and five no-stimulus catch trials. By the end of each daily session every listener had received a total of 294 target signals at four different levels, 47, 51, 53, and 55 dBA (12 in the warm-up series, 240 in the experimental series); 48 target signals at 49 dBA (18 warm-up, 30 experimental); and 15 of each of the eight probes, varying from the target signal frequency by ±25, 50, 75, and 100 Hz, all at 49 dBA (these all presented in the experimental series). Their instructions were to press the right hand ''yes'' button whenever they heard any tone, doing this as quickly as possible while trying to keep the numbers of red lights on blank trials to a minimum, and to press the left hand ''no'' button if the noise ended without their hearing a tone. The observation periods were presented at a minimum 2 s after noise offset on the preceding trial if the response occurred before noise offset; or 2 s after the response for later responses; or after a maximum of 6 s after noise offset if the listener did not respond. Only the ''yes'' response latencies were used in these analyses, and responses with a latency of less than 170 ms had both the choice and the latency discarded at the end of the experiment as being too fast to have been evoked by the signal. There were 130 ''yes'' responses in this discarded group, compared to a total of 6066 that had acceptable latencies. Each of these response latencies (RL ms) was converted to a response speed (RS$=100/$RL ms), which had less markedly skewed distributions.

III. RESULTS

A. Hearing thresholds

Figure 1 presents the thresholds in dB HL for each frequency for each of the four elderly listeners, listing their age on the right side of the figure, for the ear that was more sensitive in their particular band of the experimental stimuli (500–1000 Hz for the 82-year-old listener, 1000–1500 Hz for the others). The average threshold in this band was 9.4 dB HL (range 5–14 dB) for the better ear and 13.8 dB HL (range 9–18 dB) for the ear with less sensitivity. The most elderly participant had the greatest hearing loss, but the next oldest, at 67 years of age, had the best hearing overall. Comparison of these thresholds with the age- and gender-specific ranges provided by Morrell *et al.* (1996) indicates that all lie within the center interquartile range for age-related hearing loss. All members of the younger group had thresholds within ±10 dB HL across all frequencies, and from $+2$ to -10 dB HL within the band of their experimental stimuli. While our measure of hearing thresholds across the standard frequencies used in the audiogram does not guarantee the absence of limited spectral regions in which our research participants may have had idiosyncratic differences in sensitivity, if there were any such localized perturbations in sensitivity, they should affect both the stimulus tones and the surrounding noise and thus have no effect on the signal-to-noise ratios across listeners.

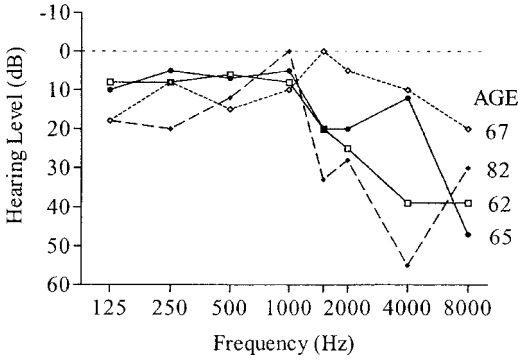


FIG. 1. Auditory thresholds (dB HL) for the ear with greater sensitivity in the region of the target signal and the probes in the four elderly listeners, and their age in years. Mean thresholds in this region ranged from 5 to 14 dB HL in the older listeners.

B. Age and signal level and the detection of frequently presented target signals

Figure 2 presents the mean detection of the target signals in the upper graph and the mean speeds in the lower graph. Both measures smoothly diminished with decreasing signal level with significant linear trends for detection probability and speeds, and an additional significant quadratic trend for probability, $p<0.0001$. Signal level accounted for 87% of the variance in detection ($F=55.19$, $df=4/28$, $p<0.0001$) and 91% of the variance in speeds ($F=77.55$, $df=4/28$, $p<0.0001$). There were no significant age effects on either measure of detection nor were there significant differences in the false alarm rate. It may be noted that the 49-dB level of the target signal provides the comparison for the probe stimuli in the subsequent analyses and here the two groups were near identical in their detection rate, at 68.2% in the young and 70.7% in the older listeners. The visual indication of an apparently stable age-related speed difference resulted because two of the five young listeners consistently responded very rapidly, while the others responded at the same mean speed as the older listeners (for the overall group difference favoring the younger group, $p=0.2$). One feature of the speed scores is their sensitivity to stimulus levels above

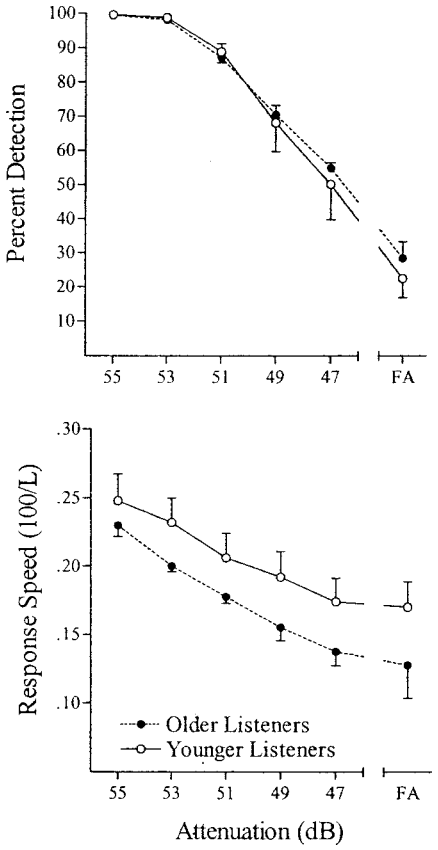


FIG. 2. Mean percent detection in the upper graph, and response speeds on successful detection trials in the lower graph (±SEM for between group comparisons) for the common signals (800 or 1200 Hz, 200-ms duration) presented at five levels, 47 to 55 dB (SPL), 2 s after the onset of a 3-s burst of 60-dB wideband noise; and false alarm detection rate and speeds for "no-stimulus" catch trials.

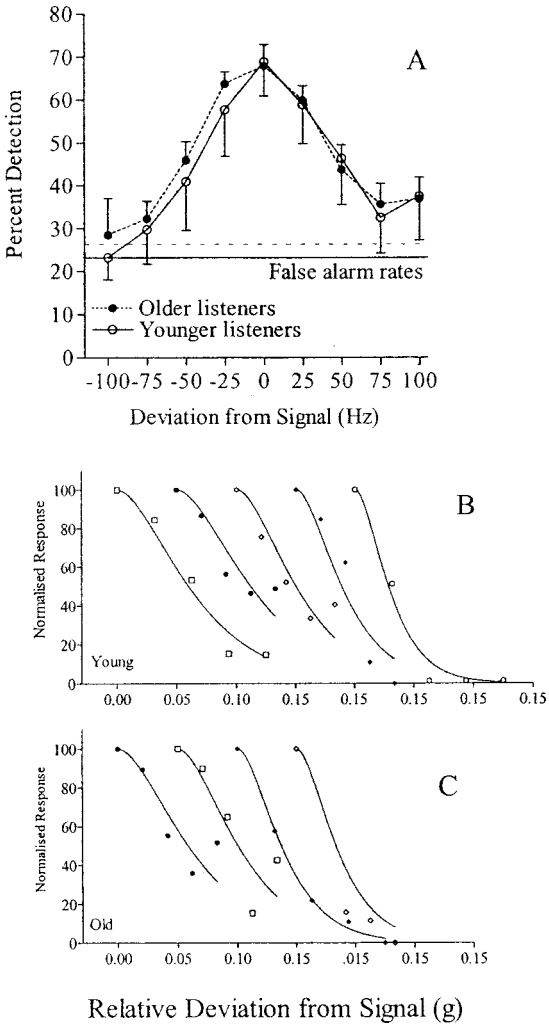


FIG. 3. (a) Mean percent detection (±SEM for between group comparisons) for the signal presented at 49 dB and for the off-frequency probes at the same level that deviated from the signal by ±25 to 100 Hz in the groups of younger listeners and older listeners, and their false alarm rates. (b) and (c) Individual normalized percent detection and one-sided listening-band filter shapes across probes of different frequencies relative to the signal [$g=(fP-fS)/fS$, see text for details], calculated from a modification of the ROEX model of Patterson *et al.* (1982): in (b), the younger listeners; in (c), the older listeners.

the point of asymptotic 100% detection as all of the listeners, young and old alike, responded more rapidly to the 55-dB signal compared to the 53-dB signal: means (SEM) = 0.240 (0.011) and 0.218 (0.011), $r=0.96$, paired $t(8)=6.86$, $p=0.0001$, and, indeed, all adjacent points were significantly different for both groups combined ($p<0.01$). This feature of the speed scores results simply because there is no significant upper limit on response speeds as there is no detection rate, and thus differences in speed across conditions can provide additional information about relative stimulus salience that is not evident in probability measures.

C. Age and tonal frequency and the detection of rarely presented probes

Figure 3(a) shows the mean probability of detection (±SEM) for the probe stimuli in the young and elderly listeners as a function of the deviation of the probes from the target signals measured in Hz, with all stimuli presented at the same 49 dB SPL level. Both groups of listeners exhibit the standard pattern of signal detection that is produced by manipulating their anticipation of the spectral location of the target signal, this yielding a systematic decline in detection as the off-frequency probe tones increasingly deviated from the target signal. In the analysis of variance the effect of spectral frequency on stimulus detection was highly significant ($F=22.72$, $df=8/56$, $p<0.0001$, $R^2=0.76$) but all effects of age had $F<1.0$. When the data were combined over both groups the adjacent means between 0-Hz deviation to 75-Hz deviation were all significantly different ($p<0.01$). The detection probabilities for each of the individual listeners is shown in Figs. 3(b) and (c). To obtain greater stability in these single-subject data the two sides of the detection patterns for each listener were averaged to obtain the one-sided means, on the assumption that the filters are symmetric in this spectral range. The x axis was transformed as a relative deviation from the signal frequency, using the formula $g=$[(probe tone frequency − target signal tone frequency)/target signal tone frequency], and the y axis was normalized so that the detection data for each listener were expressed as a relative proportion of the difference between that person's detection level of the target signal and their false alarm rate. These data are depicted offset horizontally from each other, together with one-sided listening band filters for each sub-

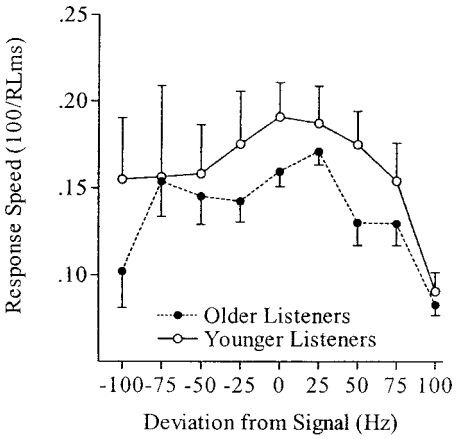


FIG. 4. Mean response speeds (±SEM for between group comparisons) on successful detection trials for the signal at 49 dB and for the off-frequency probes at the same level that deviated from the signal by ±25 to 100 Hz for the younger listeners and the older listeners.

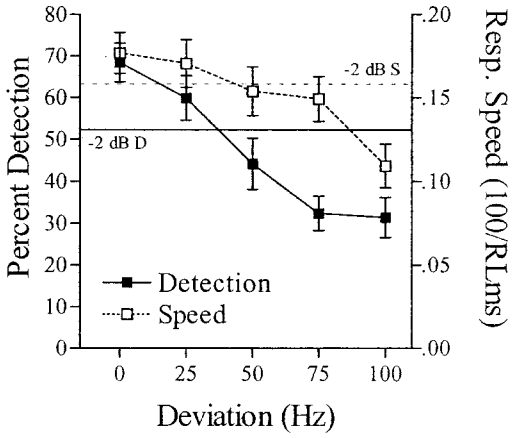


FIG. 5. Mean response speed and detection rate (±SEM) summed over the two groups of listeners and the probes at absolute deviant frequencies above and below the target signal; the horizontal lines provide the mean detection rate (solid) and the mean response speed (dotted) for the target signal tone presented at a level 2 dB below that of the probes (see Fig. 2).

ject, these derived from the formula $Y(g)=(1+pg)e^{-pg}$, following Dai *et al.* (1991). Figure 3(b) shows the individual filter shapes and the detection data for each of the five younger listeners and Fig. 3(c) shows the individual filter shapes and detection data for each of the four elderly listeners (the symbols on this plot are the same as those used to denote the different listeners in Fig. 1). The relative equivalent rectangular band (expressed as a proportion of the center frequency, rERB=ERB/target signal tone frequency) can be derived from the above formula as ERB=4 (target signal tonal frequency/p), which yields a mean rERB (±SEM) for the young listeners of 0.114±0.015, and for older listeners a mean rERB of 0.107±0.014. The fit between the calculated filter shapes and the obtained detection probabilities ranged across subjects from a low of $R^2=0.77$ to a high of $R^2=0.99$ (both values obtained in elderly listeners): the mean R^2 (SEM) for the young group was 0.91 (0.03), and for the elderly group 0.84 (0.05). Thus, rather than showing that elderly listeners have a broader ERB than the young, as had been anticipated, the difference between the groups actually favors the elderly listeners, though by a small and nonsignificant amount. The interpretation of this obtained difference is best founded in the observation that the 90% confidence interval for the obtained difference (which is a 95% CI for the hypothesis that the elderly group had the wider ERB) has an upper limit of +0.137, as this indicates that the elderly listener can be reasonably assumed to have a relative ERB that could be at most only about 20% wider than that of the younger listener. Indeed, given these data, if there were any difference in bandwidth favoring the younger listener, there is a 50% probability that it is less than 10%.

Figure 4 shows mean response speeds (±SEM) for the two groups for the trials on which the probes were detected, the *x* axis being again the simple deviation of the probe tones from the target signal in Hz. Overall it can be seen that response speeds declined with the increasing deviation of the probe from the signal, but the attentional filter for response speed depicted in Fig. 4 appears to be more variable and less sharply peaked than that for detection, as depicted in Fig. 3(a). Although the mean speeds were generally higher in the younger listeners, their standard errors overlapped with the older groups, as did the distributions of individuals within the groups: again, this difference between the groups should be attributed to the two very fast young responders, the three others providing scores that were intermingled with those of the older listeners. In the analysis of variance of these data the main effect of probe frequency yielded $F(8/56)=11.67$, $p<0.0001$, $R^2=0.59$, while no effects involving age were significant ($p>0.3$). There was no obvious difference in the shape of the response speed functions between the two groups, and although these functions are not inconsistent with a rounded exponential function, their variability precluded formal curve fitting from which an rERB could be derived. In comparing the apparently different effects of increasing deviation of the probe from the target signal on the two response measures in the contrast of Fig. 4 with Fig. 3(a), it should be appreciated that response speeds are obtained only for those probe stimuli that had been detected. It is then interesting to find that both groups responded significantly more rapidly to detected target signals than they did to detected probes when the probes deviated from the target signal by 50 Hz or more (for the two groups combined all $p<0.001$ according to Dunnett's test).

D. The relative specificity of the listening bands for detection rate and speed

Figure 5 presents the means (±SEM) for the simple nontransformed percent detection scores and the response speeds as a function of the absolute deviation in Hz of the probes from the target signal, the data combined across age groups and across the two sides of the detection pattern in order to provide the most stable data points. Mean percent detection and the mean speed scores for the target signal frequency presented at 47 dB (2 dB below that of the probes) are also presented on this graph. Although the speed function had a different shape than that of detection, being relatively more shallow for probes close to the frequency of the signal but falling off more steeply for the most disparate probe, both functions indicate that deviations between 25 and 50 Hz

away from the signal are the equivalent of a 2-dB attenuation in signal level. This 2-dB value agrees with that found by Dai *et al.* (1991, Fig. 2, p. 2839) for the detection of probes surrounding 1-kHz targets, those data obtained in the more usual 2AFC procedure. Finding similar levels of attenuation for both detection rate and speed suggests that the attentional mechanism has had its attenuating effect on the effective level of the probes even on these trials for which the probe was detected, though it is also possible that the unfamiliarity of the detected probe rather than its apparent level diminished the speed of the response.

IV. DISCUSSION

The major outcome of this study is that the functions relating detection and response speed to off-frequency probe detection in elderly listeners were nearly identical to those of a younger group, both showing the same systematic decrease in the detection of the probe as its tonal frequency deviated away from that of the more commonly presented target signal. These data reveal that the attentional filter and the frequency selectivity that it provides for simple sinusoids presented in background noise do not significantly deteriorate in the aged listener. Further, we can reasonably assert from the analysis of confidence intervals that if there is any difference at all in the relative width of the ERB favoring the younger listener, it can be at most a difference of about 20%, and there are even odds that it is less than 10%. Thus we may conclude that this form of selective auditory attention is well preserved in elderly listeners. One cautionary note is that the number of older subjects in this sample was small and homogeneous in composition, as all enjoyed the same academic background and shared a scholarly interest in the outcome of this research. This may suggest that their proficiency in the probe-signal task was not typical of the general population, but, on the other hand, their hearing thresholds for tonal stimuli across the spectrum approximated those of their age norms, and so they cannot be said to have generally superior auditory abilities.

The rationale for the present study of the effects of age on the attentional filter derived in part from the data showing that its bandwidth depends on the integrity of the olivocochlear bundle (Scharf *et al.*, 1994, 1997) and in part from other reports that the effectiveness of the olivocochlear bundle deteriorates with age (Castor *et al.*, 1994; Kim *et al.*, 2002), these together yielding the hypothesis that the attention band would be wider in older listeners. The potential importance of this work in adding to our understanding of the functional bases of presbycusis was underlined by the demonstration that deficits in OCB activity are also linked to deficits in phoneme recognition in noise (Giraud *et al.*, 1997). But then, given this empirical context, how is the present failure to find any effect of age on the attention band to be understood? And, given these negative results, must we now conclude that changes in the attention band with age do not at all contribute to the pathogenesis of presbycusis?

One clear implication of this unexpected result is that selective listening and noise suppression of otoacoustic emissions seem not in fact to be strongly related phenomena. There are several ways in which this assertion can be interpreted, some more plausible that others. The most simple hypothesis is that in contrast to the suppression of otoacoustic emissions, selective listening does not depend on OCB function after all. However, this conjecture seems to be effectively countered by the weight of the published evidence provided by Scharf and his colleagues (1994, 1997) showing that this efferent control system is essential for the acute selectivity of the attention band. But a more subtle and plausible distinction might be made between the two phenomena, for it is at least possible that they depend on different aspects of OCB function. The effect of coincident noise on the strength of otoacoustic emissions for tones is entirely due to the inhibitory effect of the crossed medial branch of the OCB on outer hair cell activity. It may be that this particular branch of the OCB or its activation at the auditory periphery is particularly sensitive to age, or it may be that otoacoustic emissions are very sensitive to just modest changes in OCB function. These hypotheses are consistent with the data provided by Kim *et al.* (2002), who found deficits in contralateral noise suppression of emissions in middle-aged as well as elderly listeners. In contrast, perhaps widening of the attention band is apparent only when the OCB is completely disrupted, as it would be following a vestibular neurotomy (Scharf *et al.*, 1994, 1997). A further complication is introduced by the fact that the crossed branch is solely responsible for the suppression of otoacoustic emissions, but both the crossed and the uncrossed branches of the OCB are severed by a vestibular neurotomy: perhaps both branches of the OCB contribute to selective attention, and perhaps its lateral uncrossed branch is less affected by age. It is also possible that while the function of the OCB evoked at the brainstem level by external noise is diminished during passive listening, its central activation is maintained in the elderly listener who is engaged in an active listening task.

There is another explanation of our failure to find an age effect, which is that typically elderly listeners with normal sensitivity for their age hear very well for predictable materials in good listening environments. We should emphasize that in the present experiment the listening conditions were optimal for focusing attention on the target signal, because its tonal frequency was highly predictable, it always occurred at a set delay time after noise onset, and this time was long enough to allow the listener to prepare for its appearance. Other studies have shown that the deleterious effects of age on perception often emerge when the test conditions are less favorable, for example, in listening to complex speech stimuli rather than simple sinusoids (Patterson *et al.*, 1982), for nonpredictable compared to predictable speech (Dubno *et al.*, 1984), and for time-compressed speech, especially in the absence of sensible context (Wingfield, 1996). The present experimental results show only that the basic neural machinery necessary for selective auditory attention under optimal conditions appears to be present and fully functional in the aged listener. Future work should seek to extend this finding, perhaps by varying the listener's uncertainty about which of several potential signals might occur in any particular trial (Macmillan and Schwartz, 1975; Schlauch and Hafter, 1991), and by providing either less preparation time between the onset of the masking noise and the signal (Dai

and Buus, 1991), or by varying the preparation time within trial blocks so as to introduce temporal as well as spectral uncertainty into the task.

Our main rationale for measuring response speed was to test a prediction that similar attention like effects should be obtained for both detection rate and response speed for probe stimuli with increasing spectral separation from the target signal, which derives from the hypothesis that the OCB attenuates the probes compared to the target signal at the level of the cochlea receptor. In fact, although the overall shapes of the two response functions were different, as could be expected because the speed scores were obtained only for probes that were detected, the early decrements in response speeds and detection were similarly affected by the change in signal frequency. This was seen in the observation that for both measures the 2-dB down points appeared at between 25 and 50 Hz away from the signal. However, for the most extreme probe, 100 Hz away from the signal, there was a very sharp decline in response speed. (There was no companion decrement in the detection of this probe compared to its neighbor, but detection was already at a near chance level.) Although this terminal difference in speeds versus detection rate may indicate some disparity between the functions relating the two measures to stimulus level, it is also possible that these response speeds were affected by the unfamiliarity of this probe when it was occasionally detected.

V. CONCLUSIONS

The present modification of the probe-signal method demonstrates that the listening or selective attention band approximates the shape of the auditory filter obtained in the typical notched noise experiments, as has been noted previously (Scharf *et al.*, 1987; Dai *et al.*, 1991), and shows also that when the infrequent probe stimuli are detected, then they are responded to relatively slowly. The surprising finding is that the specificity of the filter developed by elderly listeners in their attending to the spectral region within the masking noise that surrounds the target signal is virtually identical to that observed in younger listeners. The attention band depends on the integrity of the efferent olivocochlear bundle, and the demonstrated strength of the attentional effect in these elderly listeners apparently runs counter to the reports that the OCB is less effective in the aged. It may be that selective attention is less demanding of OCB activity than is noise-evoked suppression of otoacoustic emissions; or that the two branches of the OCB play different roles in these two phenomena; or, otherwise, that in contrast to peripheral noise evocation, central control of the auditory efferents remains robust and fully functional in the aged. However, decrements in selective attention may appear under more realistic and less optimal listening conditions, in which two or more target signals appear in the observation period with equal probability, having presentation times that vary unpredictably or provide a minimal opportunity for preparation. If deficits were to appear under these more complex conditions, it would then be appropriate to examine the relationship between selective attention and direct measures of speech recognition in difficult listening conditions in elderly listeners.

ACKNOWLEDGMENTS

This work was supported by research grants from the United States Public Health Service, AG095247, and from the Australian Research Council, 04/15/412/265. An initial report of these data was presented at the Meetings of the Association for Research in Otolaryngology, February 2001. We are grateful to two anonymous referees for their very helpful reviews of an earlier version of this report. We are especially grateful to Professor Scharf for his comments, and also for his informing us that he has seen a similar preservation of the attention band in elderly patients who were candidates for a vestibular neurotomy or had already received this treatment (see, for example, patient LG, described in Scharf *et al.*, 1997), including one patient who was 87 years of age at the time of testing.

ANSI (**1989**). ANSI S3.6-1989, ''Specifications for audiometers'' (American National Standards Institute, New York).

Buchtel, H. A., and Butter, C. M. (**1988**). ''Spatial attentional shifts: Implications for the role of polysensory mechanisms,'' Neuropsychologia **26**, 499–509.

Castor, X., Veuillet, E., Morgon, A., and Collet, L. (**1994**). ''Influence of aging on active cochlear micromechanical properties and on the medial olivocochlear system in humans,'' Hear. Res. **77**, 1–8.

Collet, L., Kemp, D. T., Veuillet, E., Duclaux, R., Moulin, A., and Morgon, A. (**1990**). ''Effect of contralateral auditory stimuli on active cochlear micromechanical properties in human subjects,'' Hear. Res. **43**, 251–261.

Dai, H., and Buus, S. (**1991**). ''Effect of gating the masker on frequency-selective listening,'' J. Acoust. Soc. Am. **89**, 1816–1818.

Dai, H., Scharf, B., and Buus, S. (**1991**). ''Effective attenuation of signals in noise under focused attention,'' J. Acoust. Soc. Am. **89**, 2837–2842.

Dubno, J. R., Dirks, D. D., and Morgan, D. E. (**1984**). ''Effects of age and mild hearing loss on speech recognition in noise,'' J. Acoust. Soc. Am. **76**, 87–96.

Giraud, A. L., Garnier, S., Micheyl, C., Lina, G., Chays, A., and Chery-Croze, S. (**1997**). ''Auditory efferents involved in speech-in-noise intelligibility,'' NeuroReport **8**, 1779–1783.

Greenberg, G. Z., and Larkin, W. (**1968**). ''Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: The probe-signal method,'' J. Acoust. Soc. Am. **44**, 1513–1523.

Houtsma, A. J. M. (**1995**). ''Pitch perception,'' Chap. 8 in *Hearing*, edited by B. C. J. Moore (Academic, San Diego) pp. 267–295.

James, W. (**1890**). *The Principles of Psychology* (Holt, New York).

Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Kim, S. H., Frisina, D. R., and Frisina, R. D. (**2002**). ''Effects of age on contralateral suppression of distortion-product otoacoustic emissions in human listeners with normal hearing,'' Audiol. Neuro-Otol. (in press).

Liberman, M. C. (**1989**). ''Rapid assessment of sound-evoked olivocochlear feedback: suppression of compound action potentials by contralateral sound,'' Hear. Res. **38**, 47–56.

Macmillan, N. A., and Schwartz, M. (**1975**). ''A probe-signal investigation of uncertain-frequency detection,'' J. Acoust. Soc. Am. **58**, 1051–1058.

Mondor, T. A., and Zatorre, R. J. (**1995**). ''Shifting and focusing auditory spatial attention,'' J. Exp. Psychol. Hum. Percept. Perform. **21**, 387–409.

Moore, B. C. J. (**1973**). ''Frequency difference limens for short-duration tones,'' J. Acoust. Soc. Am. **54**, 610–619.

Morrell, C. H., Gordon-Salant, S., Pearson, J. D., Brant, L. J., and Fozard, J. L. (**1996**). ''Age- and gender-specific reference ranges for hearing level and longitudinal changes in hearing level,'' J. Acoust. Soc. Am. **100**, 1949–1967.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). ''The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold,'' J. Acoust. Soc. Am. **72**, 1788–1802.

Posner, M. I. (**1978**). *Chronometric Explorations of Mind* (Erlbaum, Hillsdale, NJ).

Scharf, B. (**1998**). ''Auditory attention: The psychoacoustical approach,'' Chap. 2 in *Attention*, edited by H. Pashler (Psychology, Hove, East Sussex, UK), pp. 75–117.

Scharf, B., Magnan, J., and Chays, A. (**1997**). ''On the role of the olivocochlear bundle in hearing: Sixteen case studies,'' Hear. Res. **103**, 101–122.
Scharf, B., Magnan, J., Collet, L., Ulmer, E., and Chays, A. (**1994**). ''On the role of the olivocochlear bundle in hearing: A case study,'' Hear. Res. **75**, 11–26.
Scharf, B., Quigley, S., Aoki, C., Peachey, N., and Reeves, A. (**1987**). ''Focused auditory attention and frequency selectivity,'' Percept. Psychophys. **42**, 215–223.
Schlauch, R. S., and Hafter, E. R. (**1991**). ''Listening bandwidths and frequency uncertainty in pure-tone signal detection,'' J. Acoust. Soc. Am. **90**, 1332–1339.
Spence, C. J., and Driver, J. (**1994**). ''Covert spatial processing in audition: Exogenous and endogenous mechanisms,'' J. Exp. Psychol. Hum. Percept. Perform. **20**, 555–574.
Spence, C. J., and Driver, J. (**1996**). ''Audiovisual links in endogenous covert spatial attention,'' J. Exp. Psychol. Hum. Percept. Perform. **22**, 1005–1030.
Swets, J. A., and Kristofferson, A. B. (**1970**). ''Attention,'' Annu. Rev. Psychol. **21**, 339–366.
Swets, J. A., Shipley, E. F., McKey, M. J., and Green, D. M. (**1959**). ''Multiple observations of signals in noise,'' J. Acoust. Soc. Am. **31**, 514–521.
Tanner, W. P., Jr., Swets, J. A., and Green, D. M. (**1956**). ''Some general properties of the hearing mechanism,'' Technical Report No. 30, Electronic Defense Group, University of Michigan, Ann Arbor, MI.
von Bekesy, G. (**1960/1929**). *Experiments in Hearing*, translated and edited by E. G. Wever (McGraw-Hill, New York).
Warr, W. B., Guinan, J. J., Jr., and White, J. S. (**1986**). ''Organization of the efferent fibers: The lateral and medial olivocochlear systems,'' Chap. 18 in *Neurobiology of Hearing: The Cochlea*, edited by R. A. Altschuler, D. W. Hoffman, and R. P. Robbin (Raven, New York), pp. 333–348.
Wiederhold, M. L. (**1986**). ''Physiology of the olivocochlear system,'' Chap. 19 in *Neurobiology of Hearing: The cochlea*, edited by R. A. Altschuler, D. W. Hoffman, and R. P. Robbin (Raven, New York), pp. 349–370.
Wingfield, A. (**1996**). ''Cognitive factors in auditory performance: context, speed of processing, and constraints of memory,'' J. Am. Acad. Audiol **7**, 175–182.
Woodworth, R. S., and Schlosberg, H. (**1955**). *Experimental Psychology* (Holt, New York).
Working Group on Speech Understanding and Aging (**1988**). ''Speech understanding and aging,'' J. Acoust. Soc. Am. **83**, 859–895.

Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses

Li Xu[a)]
Kresge Hearing Research Institute, Department of Otolaryngology, University of Michigan, 1301 East Ann Street, Ann Arbor, Michigan 48109 and School of Hearing, Speech and Language Sciences, Ohio University, Grover Center, Athens, Ohio 45701

Yuhjung Tsai and Bryan E. Pfingst
Kresge Hearing Research Institute, Department of Otolaryngology, University of Michigan, 1301 East Ann Street, Ann Arbor, Michigan 48109

(Received 12 July 2001; revised 23 January 2002; accepted 25 April 2002)

Tone languages differ from English in that the pitch pattern of a single-syllable word conveys lexical meaning. In the present study, dependence of tonal-speech perception on features of the stimulation was examined using an acoustic simulation of a CIS-type speech-processing strategy for cochlear prostheses. Contributions of spectral features of the speech signals were assessed by varying the number of filter bands, while contributions of temporal envelope features were assessed by varying the low-pass cutoff frequency used for extracting the amplitude envelopes. Ten normal-hearing native Mandarin Chinese speakers were tested. When the low-pass cutoff frequency was fixed at 512 Hz, consonant, vowel, and sentence recognition improved as a function of the number of channels and reached plateau at 4 to 6 channels. Subjective judgments of sound quality continued to improve as the number of channels increased to 12, the highest number tested. Tone recognition, i.e., recognition of the four Mandarin tone patterns, depended on both the number of channels and the low-pass cutoff frequency. The trade-off between the temporal and spectral cues for tone recognition indicates that temporal cues can compensate for diminished spectral cues for tone recognition and vice versa. An additional tone recognition experiment using syllables of equal duration showed a marked decrease in performance, indicating that duration cues contribute to tone recognition. A third experiment showed that recognition of processed FM patterns that mimic Mandarin tone patterns was poor when temporal envelope and duration cues were removed. [DOI: 10.1121/1.1487843]

PACS numbers: 43.66.Ts, 43.71.Hw, 43.71.Es, 43.71.Gv [MRL]

I. INTRODUCTION

More than a quarter of the world's population speak one of the tone languages, which include Mandarin Chinese, Cantonese, Vietnamese, Thai, etc. In fact, Mandarin Chinese is spoken by more people than any other language. In Mandarin Chinese, there are four tone patterns as defined by the fundamental frequency of voiced speech. Other features that vary with the tone patterns include amplitude pattern and syllable duration. In isolated syllables, the patterns of the fundamental frequency for tone patterns, referred to as tone 1, 2, 3, and 4 in the following, are (1) flat and high; (2) rising; (3) falling and then rising; and (4) falling [Fig. 1(A)]. The distinguishing characteristic of the tone language is that the variations of the tone patterns of each syllable convey lexical meaning. For example, tones 1 through 4 associated with the Chinese syllable /xu/ (pronounced as ''shoo'') can mean (1) ''void;'' (2) ''slowly;'' (3) ''permit;'' and (4) ''sequence;'' respectively.

The contemporary cochlear implant has provided benefit to thousands of patients with profound hearing loss throughout the world. However, for socioeconomic reasons, only a relatively small number of deaf people in China and Southeast Asia, where tone languages are most prevalent, have received the commercially available cochlear implants. Sporadic reports with a limited number of native Chinese speaking patients have shown that those patients, using the multichannel cochlear implants, have achieved tone recognition at various levels. Tone recognition varies from around chance performance to nearly 100% correct levels (Xu *et al.*, 1987; Zeng, 1995; Huang *et al.*, 1996; Sun *et al.*, 1998; Wei *et al.*, 2000).

What features of the tone languages are important for tonal-speech recognition? The answer to this question is crucial for understanding the mechanism by which the implant patients perceive the tone patterns of tone languages and for designing future cochlear implants that are effective for patients who speak tone languages. In the current speech processing strategies of cochlear implants, the speech signal is typically represented in electrical pulses that are amplitude modulated by the envelope of the signal. In the CIS (continuous interleaved sampling) strategy (Wilson *et al.*, 1991), the speech signal is divided into a number of bands and the envelope of each band is extracted and used to modulate the pulse trains. In the SPEAK (spectral peak) or ACE (advanced combined encoding) strategies (Skinner *et al.*, 1994; Seligman and McDermott, 1995; Cochlear Corporation,

a)Author to whom correspondence should be addressed. Electronic mail: XuL@ohio.edu

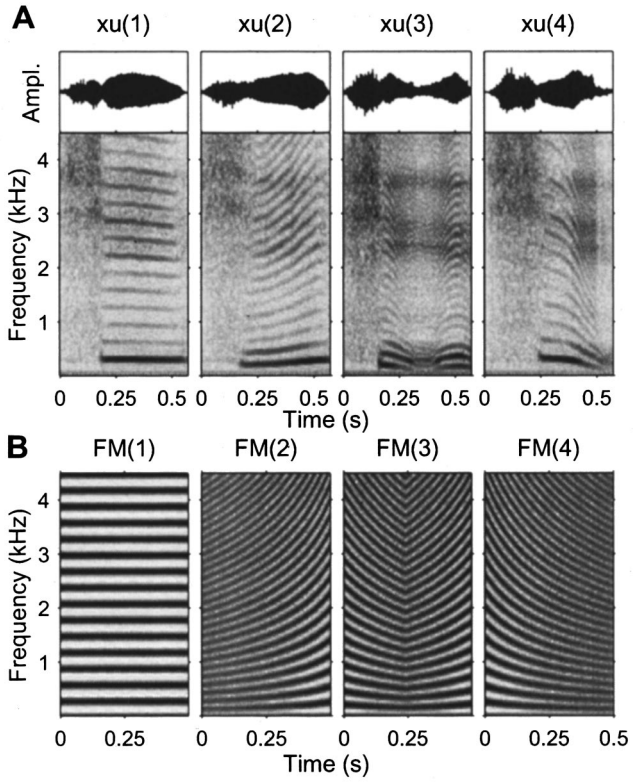


FIG. 1. Acoustic features of the speech and artificial signals. (A) Raw waveforms (top row) and spectrograms (bottom row) of the Chinese syllable /xu/ (pronounced ''shoo'') spoken by a female. Panels from left to right show tone patterns 1 through 4. The lexical meaning associated with tones 1 through 4 are ''void,'' ''slowly,'' ''permit,'' and ''sequence,'' respectively. The darkness of the spectrograms represents the energy associated with time and frequency. The fundamental frequency and the harmonics of the voiced part (/u/) show flat, rising, falling/rising, and falling patterns for tones 1 through 4, respectively. The temporal envelopes of the waveforms also differ from one tone pattern to another. The durations of the syllables are about 0.6 s with the voiced part averaged around 0.4 s. (B) Spectrograms of the higher-pitched set of the frequency-modulated (FM) sweeps synthesized to mimic the four tone patterns of Mandarin Chinese. The fundamental frequencies are listed in Table I. The durations are constant at 0.5 s.

1999), the speech signal is sent to a bank of filters (20 for SPEAK and 22 for ACE) and the envelopes of five to ten of the largest outputs of the 20 or 22 filters are extracted. In a recent study, Fu and colleagues (1998b) simulated one- through four-channel CIS-type cochlear prosthesis and tested speech recognition in normal-hearing native Chinese speakers. Their results showed that the recognition scores of Chinese vowels, consonants, and sentences bore a similar pattern to that seen in English (Shannon *et al.*, 1995). They also found that tone recognition scores were high and largely independent of the number of filter bands. In contrast, Wei and colleagues (2001) showed in a preliminary report that the tone-recognition performance in a group of four cochlear implant users depended on the number of channels and reached a plateau at about ten channels. These apparently conflicting results motivated the following questions: (1) what would the tonal-speech recognition performance be with more than four channels; (2) how would the subjects judge the sound quality of the acoustic simulations of cochlear prosthesis; and (3) what features of the speech signals provide the robust cues for tone recognition under cochlear prosthesis-type stimulation?

The present study was designed to address those questions. Using the methods of acoustic simulations of cochlear prosthesis as developed and used by Shannon and colleagues (1995, 1998) and other researchers (Dorman *et al.*, 1997a, 1997b), we examined the features of stimulation that affect tonal-speech recognition and subjective sound quality judgments. The effect of spectral resolution was studied by varying the number of simulated channels from 1 to 12. The temporal features of the speech were studied by varying the low-pass cutoff frequencies used in extracting the amplitude envelopes. According to Rosen (1992), the temporal features of speech signals can be partitioned into three categories, i.e., envelope (2–50 Hz), periodicity (50–500 Hz), and fine structure (500–10 000 Hz). In the present study we modified the envelope and periodicity cues as the temporal features for speech recognition of Mandarin Chinese. We varied the low-pass cutoff frequencies systematically from 1 to 512 Hz in 1-octave steps, to examine the details of the temporal features of stimulation that might contribute to the tone recognition. The effects on tone recognition of the interaction between the temporal and spectral features of stimulation were then studied. In order to gain control of the syllable duration and overall amplitude (i.e., suprasegmental features), which might convey information for tone recognition, new sets of speech signals that were carefully matched in duration were recorded. Tone recognition on those equal-duration syllables was tested at randomly varying sound levels. In addition, frequency-modulated (FM) sweeps that mimicked the Mandarin Chinese tone patterns were generated [Fig. 1(B)] so as to gain parametric control of the temporal and spectral features of the stimuli. Recognition of the FM sweeps processed through the cochlear prosthesis simulation models was then tested.

II. METHOD

A. Subjects

Ten normal-hearing native Mandarin Chinese speakers were recruited from the student and staff population at the University of Michigan. All subjects had pure-tone air-conduction thresholds$\leqslant$20 dB HL at octave frequencies between 250 and 8000 Hz. Subjects ranged from 29 to 43 years of age (35.6 ± 4.4, mean and s.d.). There were eight females and two males. All subjects had at least college-level education in China. The use of human subjects in this study was reviewed and approved by the University of Michigan Medical School Institutional Review Board.

B. Speech test materials and frequency-modulation sweeps

The Mandarin Chinese speech test materials consisted of consonant, vowel, tone, and sentence tests in both male and female voices. All of the combinations of consonant–vowel–tone syllables used in this study were real words in Chinese. The tone, consonant, and vowel tests were in a four-choice format in which the subjects were required to choose one of the selections from a list of four words. The consonant and

TABLE I. The fundamental frequencies (Hz) of the low- and high-pitched sets of four patterns of FM sweeps. The four patterns are (1) flat and high; (2) rising; (3) falling and then rising; and (4) falling, thus mimicking the tone patterns of Mandarin Chinese. Listed in the table are the initial and final frequencies of the FM sweeps. For FM pattern 3, the frequency changes direction in the middle of the sweep and that frequency is also listed in the table.

	Initial		Middle		Final	
Patterns	Low	High	Low	High	Low	High
(1) –	150	300	···	···	150	300
(2) /	75	150	···	···	150	300
(3) V	125	250	89	178	125	250
(4) \	150	300	···	···	75	150

vowel lists of the male and female voices were compiled from the Speech Audiometry Tables developed by the Department of Otolaryngology at the PLA General Hospital, Beijing (Gu, 1988). There were 40 lists of four Chinese monosyllable words for consonant and vowel tests. In each list for the consonant test, the four syllables had the same vowel and tone, e.g., du(2), mu(2), ru(2), and shu(2). In each list for the vowel test, the four syllables had the same consonant and tone, e.g., da(4), dai(4), di(4), and du(4). The tone and sentence tests of the male voice were initially developed by Beijing Union Hospital for audiometric test of patients (Zhang *et al.*, 1988) and were further compiled by Fu *et al.* (1998b). We recorded the female voice counterpart of the tone and sentence test materials. There were ten lists of four syllables for the tone tests and each syllable in a given list had the same consonant and vowel, e.g., ma(1), ma(2), ma(3), and ma(4). For each tone test, the ten lists were administered four times for a total of 40 lists. In these 40 lists, the male voice appeared in half of the cases and the female voice appeared in the other half, in random order. The sentence test materials consisted of ten lists of ten short sentences, each of which had four to seven key words. Each sentence test used one of the sentence lists. Additionally, we recorded 50 different short sentences used for practice purposes only. All speech test materials were digitized at a sampling rate of 22 050 Hz and stored in a 16-bit format.

The duration of naturally spoken monosyllable words varied systematically with the tone patterns, with tone 3 being the longest and tone 4 the shortest (Whalen and Xu, 1992; also see Fig. 8). This duration cue might contribute to the recognition of the tones. To eliminate this potential cue, we conducted an additional experiment in which we rerecorded the ten lists of the tone test materials multiple times using both male and female voices. The waveforms of the multiple recordings were then visualized with sound-analysis software [COOLEDIT 2000 (Syntrillium Software, Scottsdale, AZ)]. Tokens in which the durations of all four tones of the same syllable were within 1-ms precision were selected. These equal-duration speech materials were then used for additional tone recognition tests as described below.

To gain better parametric control of the stimuli, two sets of four frequency-modulation (FM) sweeps that mimicked the four Mandarin Chinese tone patterns were generated for recognition tests as described below. All FM sweeps were pulse trains of 500 ms in duration. Table I lists the initial, middle, and final fundamental frequencies ($F0$) of the two sets of four patterns of FM sweeps. The sampling rate was of 30 kHz. The $F0$ of the higher-pitched FM sweeps were 1 octave above the lower-pitched FM sweeps. The $F0$ of the two sets of FM sweeps were chosen to represent the upper range of the $F0$ of a male and a female voice (Fletcher, 1953; Lehiste, 1970). Different patterns and sets of FM sweeps were achieved by varying the interpulse intervals of the pulse train. Frequency analysis of the FM sweeps confirmed that they contained the $F0$ and all the high harmonics of equal energy up to the Nyquist frequency (15 kHz). Figure 1(B) shows the spectrograms of the higher-pitched FM patterns.

C. Signal processing

Signal processing for acoustic simulations of cochlear prosthesis was performed in MATLAB (MathWorks, Natick, MA). Mandarin Chinese speech signals were first passed through a bank of analysis filters (sixth-order elliptic bandpass filters). To study the effects of number of channels on recognition of tones, consonants, vowels, and sentences, the number of analysis filters was varied between 1 and 12 (1, 2, 3, 4, 6, 8, 10, and 12). The overall bandwidth was from 150 to 5500 Hz. The bandwidth of each analysis filter was based on estimated equal distance along the basilar membrane of the cochlea. The corner frequencies (F) were calculated with Greenwood's (1990) formula, $F=165.4(10^{0.06x}-1)$, where x is the distance in mm from the apex. For instance, the corner frequencies for the 6-channel analysis filters were 150, 345, 660, 1171, 1997, 3335, and 5500 Hz. Next, the temporal envelope of each analysis band was extracted by half-wave rectification and low-pass filtering (second-order Butterworth filters). The low-pass cutoff frequencies (LPFs) of the low-pass filters were fixed at 512 Hz for speech materials of consonants, vowels, and sentences, but were varied between 1 and 512 Hz in 1-octave steps for the tones in order to study the effects of the temporal features on the tone recognition. The temporal envelope of each band was then used to modulate a white noise that was bandpassed through the same analysis filter. Finally, the modulated noise bands were summed and the resultant signals were stored on the computer hard disk for presentation.

For the FM sweep recognition, FM signals were processed through the same cochlear prosthesis simulation models with number of channels varying from 4 to 40 (4, 6, 8, 10, 12, 16, 20, 30, and 40), and the LPFs varying in 2-octave

steps from 4 to 1024 Hz (4, 16, 64, 256, and 1024 Hz). In pilot experiments, we found that the higher numbers of channels and higher LPFs were required to achieve recognition more than twice the chance performance with the FM sweeps.

D. Procedures

The speech signals were presented through a loudspeaker positioned 1 m away from the subject at 0° azimuth inside an Acoustic Systems (model RE2 242S) double-walled sound-attenuating booth. The subjects were free to adjust the sound volume to the most comfortable levels; levels typically measured ~60 dB (A) with a sound level meter (Brüel and Kjær type 2231). A graphical user interface (GUI) was built to present the tone, consonant, and vowel tests in the four-choice format. In the GUI, four buttons labeled with *pinyin* (i.e., phonemic spellings) and Chinese characters were shown on a computer screen. After listening to the acoustic presentation of each speech signal, the subjects were required to indicate what they had heard by pointing and clicking on one of the four buttons using a computer mouse. After each response, the GUI would refresh the screen, display a new list of four choices, and present the next stimulus. A tone, consonant, or vowel test consisted of 40 four-choice trials. The sentence test consisted of ten sentences. The subjects were required to write down each sentence that they had heard. A different GUI was built to record the subjective judgments of the sound quality. In this GUI, a sliding bar was plotted from the left side of the screen to the right side of the screen, representing poor to excellent sound quality. The subjects were instructed to slide a marker, using a computer mouse, to a location that represented their judgment of the sound quality of the speech signals (''very poor'' to ''excellent''). For the FM pattern recognition test, the GUI showed four buttons labeled with the four respective patterns of the FM [(1) –, (2) /, (3) V, and (4) \].

All subjects received 4 to 5 h of training in listening to the processed speech materials. The speech materials for training consisted of 20 sentences randomly selected from the 50 practice sentences for all the testing conditions (i.e., number of channels) starting with 12 channels and ending with 1 channel. Subjects reported verbally what they heard. Feedback was provided during practice. Additional brief training was provided for consonant, vowel, and tone recognition tests to familiarize the subjects with the task and the use of a computer mouse.

The experiments were divided into four sessions. In the first session, subjects were tested on consonant, vowel, tone, and sentence recognition of speech materials that were processed through 1 to 12 channels with the LPF fixed at 512 Hz. The testing order for the number of channels was randomized. Each subject performed two speech tests that used both male and female voices that were presented separately. Thus, a total of 160 items was tested for consonant, vowel, and tone recognition and 40 sentences (about 200 words) were tested for sentence recognition for each number of channels. After each set of consonant, vowel, tone, and sentence tests, a subjective judgment of the sound quality was obtained. One of the subjects (Subject 1) was the female speaker of the sentence materials. Therefore, she was not tested on sentence recognition and no subjective judgments of the sound quality were obtained from her.

In the second session, subjects were tested for tone recognition only. The signals were processed through 1 to 12 analysis channels with the LPFs varied from 1 to 512 Hz in 1-octave steps. Subjects were tested twice on male voice and twice on female voice (a total of 160 items) for each combination of number of channels and LPFs. Subject 6 could not participate in this experiment session.

The third session was similar to the second session except that the speech signals were of equal durations for the four tone patterns. Also, the level of each presentation was randomized among 55, 60, 65, and 70 dB (A) to confound possible cues derived from the overall amplitude. Four of the ten subjects (Subject 2, 3, 7, and 8) participated in this session of the experiment.

The fourth session tested the recognition of the four FM sweeps that were processed through the cochlear prosthesis model. The subjects were trained for about half an hour on the unprocessed and processed FM sweeps. A total of 7200 (i.e., $80\times9\times5\times2$) responses was collected from each subject to 80 presentations of each combination of number of channels (9) and LPFs (5) and pitch (2). The order of stimuli was randomized. The level of each presentation was also randomized as in the third session. The same four of the ten subjects (Subject 2, 3, 7, and 8) who participated in the third session participated in this session of the experiment.

E. Data analysis

Arcsine-transformed percent-correct data from the speech recognition tests were used to compare the speech recognition performance under various test conditions. The arcsine transformation takes the form of $y=2\times\sin^{-1}\sqrt{p}$, where p is a percent-correct score. The rationale for arcsine transformation is that the data of percent-correct have nonuniform variance, whereas the transformed data have the property of stabilized variance of binomial data and thus are more suitable for analysis of variance (ANOVA) and other statistical analysis (Thornton and Raffin, 1978; Studebaker, 1985). A one-way ANOVA was employed to determine whether the effects on speech recognition of the independent variables (number of channels or LPFs) were statistically significant. A two-way ANOVA was used to determine whether tone recognition scores were affected by the number of channels or LPFs or both. When the ANOVA revealed a significant difference, the Tukey test was used for *post hoc* pairwise comparisons of the speech-recognition scores among the dependent variables such as number of channels or LPFs. The Tukey test was chosen because it provides a good control of the increased cumulative type-I errors inherent in repeated pairwise comparison (Keppel, 1991).

To study the tone recognition based on the syllable duration alone, a maximum-likelihood estimator was used as an ideal observer (Green and Swets, 1966). In that procedure, four histograms were generated for the durations of four tone patterns using a time bin of 60 ms. The number of syllables that fell into each time bin was then compared across the

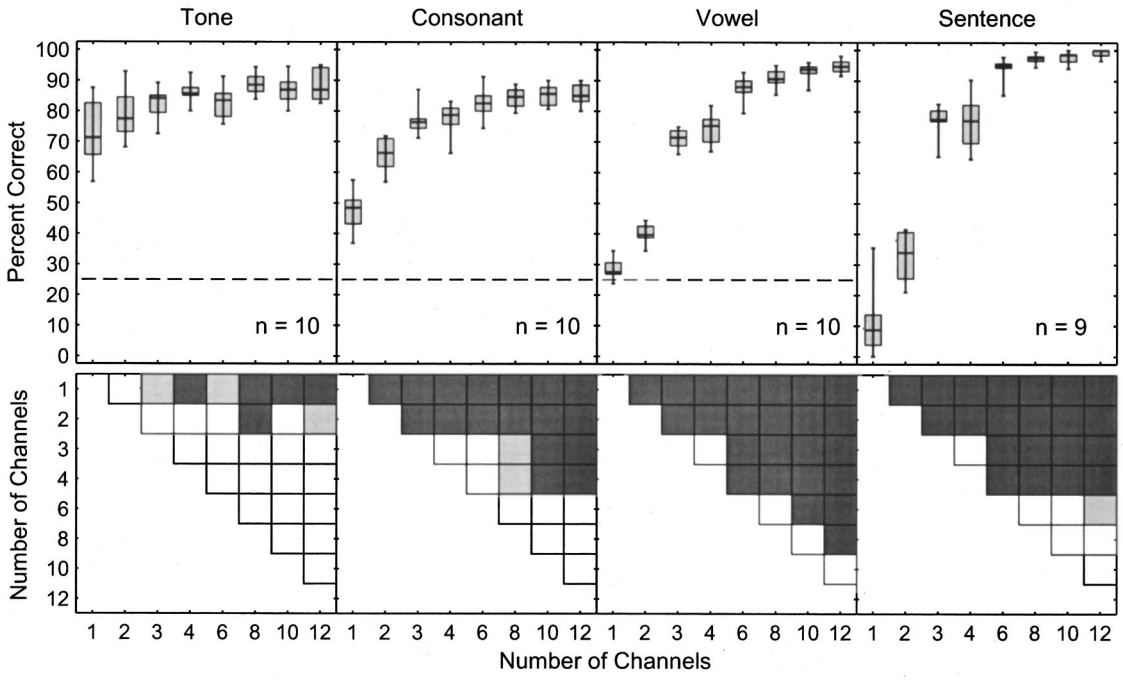


FIG. 2. Tone, consonant, vowel, and sentence recognition as a function of number of channels. The upper panels plot the distribution of the percent-correct scores across all subjects in a boxplot format in which the three horizontal lines represent the 25th, 50th, and 75th percentiles and the ends of the vertical bars represent the minimum and maximum. Panels from left to right are for tone, consonant, vowel, and sentence recognition, respectively. The dashed line at 25% for tone, consonant, and vowel recognition indicates the chance performance. For sentences, chance performance was 0%. The number of subjects tested is indicated in the lower right corner of each panel. The lower panels show the statistical significance of pairwise comparison of the mean percent correct associated with number of channels as revealed by the Tukey test. The light- and dark-gray squares represent the significance levels at $p<0.05$ and $p<0.01$, respectively. The empty squares represent comparisons not statistically significant.

four histograms. The maximum numbers of all of the time bins were then summed and the sum was divided by the total number of syllables to derive a percent-correct value for the ideal observer.

III. RESULTS

We found no significant differences between speech recognition scores for male versus female voices (t test, $p>0.05$). The frequency analysis of the male and female voices showed that the fundamental frequencies for the flat tone pattern (tone 1) were approximately 150 and 300 Hz, respectively. It seemed that the periodicity features did not contribute significantly to Mandarin Chinese speech recognition using the cochlear prosthesis acoustic simulations or that the effects of the periodicity features were confounded by other features in the speech signal. In the following report, the speech recognition scores of both male and female voices were combined.

A. Dependence on number of channels

Figure 2 (upper panels) summarizes performance in tone, consonant, vowel, and sentence recognition as a function of number of channels. The speech recognition scores (upper panels) increased as the number of channels increased from one to six and plateaued at the number of channels greater than or equal to six. For number of channels below six, the strength of dependence on number of channels was larger for vowel recognition than for consonant recognition. The strength of dependence on number of channels was the weakest for tone recognition. The dependence of sentence recognition on the number of channels between one and six was fairly strong, although this dependence was not readily comparable to the dependence on number of channels for tone, consonant, and vowel recognition because the tests had different chance levels to start with. It is noteworthy that the mean performance of sentence recognition with only one channel was a little above 10% correct, consistent with the results of the Fu *et al.* (1998b) study. One of our subjects reached a surprisingly high score of 35% correct. For native English speakers, the one-channel acoustic simulations resulted in an average of only 3% correct for English sentence recognition (Shannon *et al.*, 1995). The apparent differences between Mandarin Chinese and English suggest that the high-level tone recognition of Mandarin Chinese might contribute to the relatively higher sentence-recognition scores under the situations of one-channel stimulation.

The ANOVA revealed that the mean speech recognition scores were significantly different across numbers of channels ($p<0.01$). Pairwise comparisons between the various numbers of channels were then carried out for speech recognition scores. The results are shown in the lower panels of Fig. 2, in which each small square represents one comparison, and light- and dark-gray-filled squares represent significance levels of $p<0.05$ and $p<0.01$ (the Tukey test), respectively. The tone-recognition scores with the number of channels greater than or equal to three were significantly higher than those with only one channel. Also, the tone-recognition scores with 8 and 12 channels were significantly higher than those with two channels. Beside those differences, a larger number of channels in general did not result

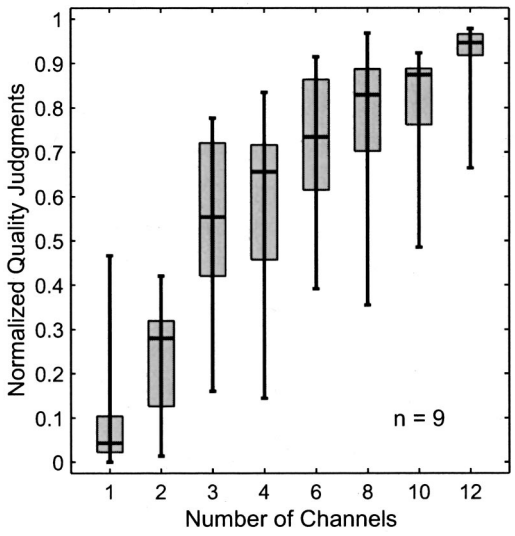


FIG. 3. Subjective judgments of the sound quality as a function of number of channels. The subjective judgments of each subject were normalized to his or her highest judgment score across all tests. The boxplot shows the distribution of the mean normalized quality judgments of all nine subjects. In the boxplot, the three horizontal lines represent the 25th, 50th, and 75th percentiles, and the ends of the vertical line show the minimum and the maximum of the distribution.

in significantly higher scores for tone recognition. The consonant-, vowel-, and sentence-recognition scores increased as a function of number of channels from one through six. While no further improvement was evident for consonant and sentence recognition with number of channels beyond six, vowel recognition did show small but significant improvement with a larger number of channels.

B. Subjective judgments

Despite the fact that the speech-recognition scores generally reached a plateau at six channels, the subjective judgments of sound quality continued to favor a larger number of channels. Figure 3 plots the distribution of mean quality judgments as a function of number of channels obtained from nine of the ten subjects. The range of quality judgments was highly variable from subject to subject. The quality judgments across all speech tests ranged from 0.10 to 0.99 for one subject to 0.02 to 0.64 for another subject. To facilitate comparison across subjects, all quality judgments from each subject were normalized to his/her maximum quality judgment of all tests. The normalized quality judgments of each condition (i.e., number of channels) were then averaged. Among subjects, the normalized quality judgments showed large variability. A side effect of this normalization procedure was that it minimized the differences in absolute levels of quality judgments among subjects. Nevertheless, there was a general trend that subjective sound quality improved as the number of channels increased up to 12.

Figure 4 plots the relationships between the speech recognition scores for tones, consonants, vowels, and sentences and the normalized quality judgments. Each dot represents results from one administration of the speech tests for one subject. The correlation coefficients (r) were fairly high for the vowel and sentence perception, but low or moderate for the tone and consonant perception. It was not surprising to see that the tone recognition scores and the quality judgments showed weak correlation because the tone recognition exhibited little dependence on number of channels (Fig. 2), whereas the subjective judgments of sound quality generally depended on the number of channels (Fig. 3). This result was also in agreement with Lehiste's (1970) observations in other tone languages that subjective judgments of phonetic quality did not seem to depend on tone recognition. It was interesting that the vowel and sentence recognition correlated more strongly with the subjective judgment of sound quality than did the consonant recognition. Perhaps the voicing of the speech contributed more to the subjective judgments than did the unvoiced part of the speech.

C. Dependence of tone recognition on the temporal features of stimulation

The low-pass cutoff frequencies (LPFs) of the low-pass filters used in extracting the amplitude envelopes determine the amount of temporal detail in the envelope of the signal

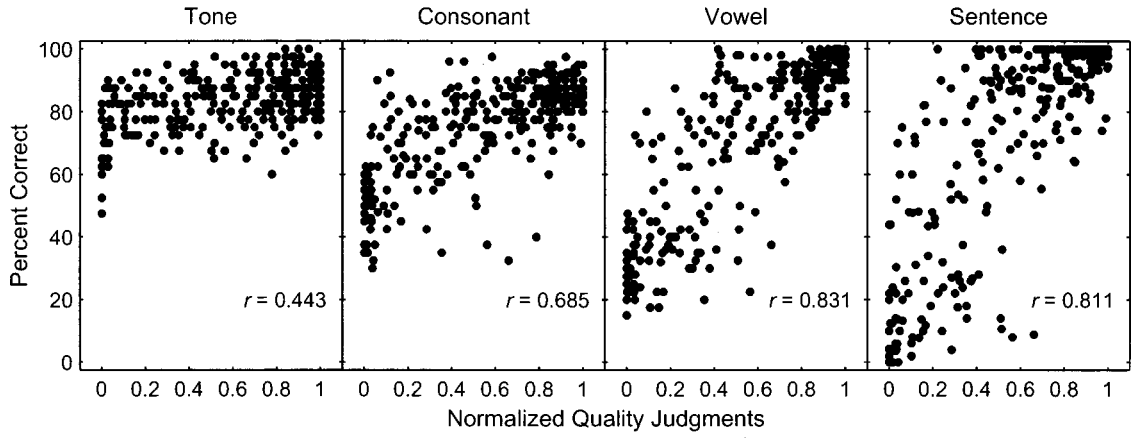


FIG. 4. Relationship between the subjective judgments of sound quality and the tone, consonant, vowel, and sentence recognition scores. Each dot represents percent correct from one speech test (ordinate) and the normalized quality judgment (abscissa). The correlation coefficients (r) are shown in the lower right corners of all panels.

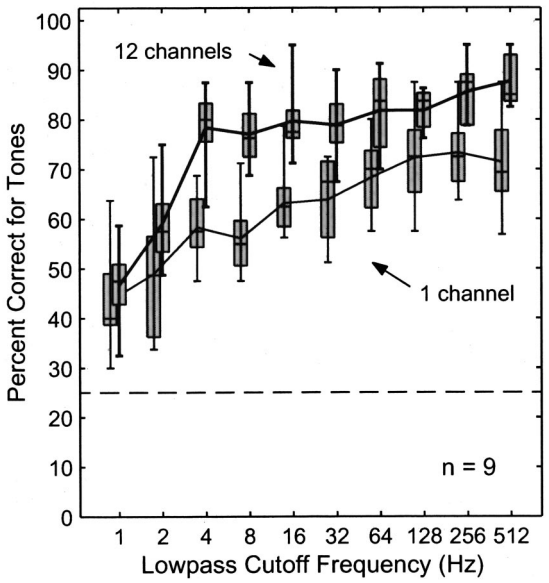


FIG. 5. Distribution of tone recognition scores as a function of the low-pass cutoff frequency. In the boxplot, the three horizontal lines represent the 25th, 50th, and 75th percentiles and the ends of the vertical line show the minimum and the maximum of the distribution across all nine subjects. The group means are connected by the solid lines. The dashed line represents the chance performance at 25%. The upper and lower traces represent data obtained with 12 channels and 1 channel, respectively, as indicated by the labels.

in each channel. Fu and colleagues (1998b) tested the tone recognition at LPFs of 50 and 500 Hz for one- through four-channel processors and found that the tone recognition scores were around 65% and 80%, respectively. In the present study, we examined the dependence of tone recognition on the LPFs that ranged from 1 to 512 Hz in 1-octave steps. Figure 5 shows the distribution of the pooled mean percent correct for tone recognition using 1 or 12 channels. Using 12 channels (Fig. 5, upper trace) typically resulted in better performance of tone recognition than using only 1 channel (Fig. 5, lower trace), as shown also in Fig. 2. The increase in the tone recognition scores as a function of LPFs was consistent. The ANOVA showed highly significant differences in the tone recognition scores among LPFs for all the channel numbers ($p<0.01$).

D. Interaction between spectral and temporal cues for tone recognition

Tone recognition was tested in nine subjects on all combinations of eight different numbers of channels and ten low-pass cutoff frequencies (LPFs) of the filter used for the envelope extraction. Figure 6 shows the percent-correct results for tone recognition from three individual subjects. Subject 3 [Fig. 6(A)] had relatively low tone recognition scores compared to other subjects. In contrast, Subject 5 [Fig. 6(C)] had relatively high scores. Subject 4 [Fig. 6(B)], however, was representative of most subjects. Most subjects showed high tone recognition scores with larger number of channels and/or with higher LPFs. There was a gradient of performance along the main diagonal line in plots like Fig. 6. This illustrated a trade-off between the number of channels and

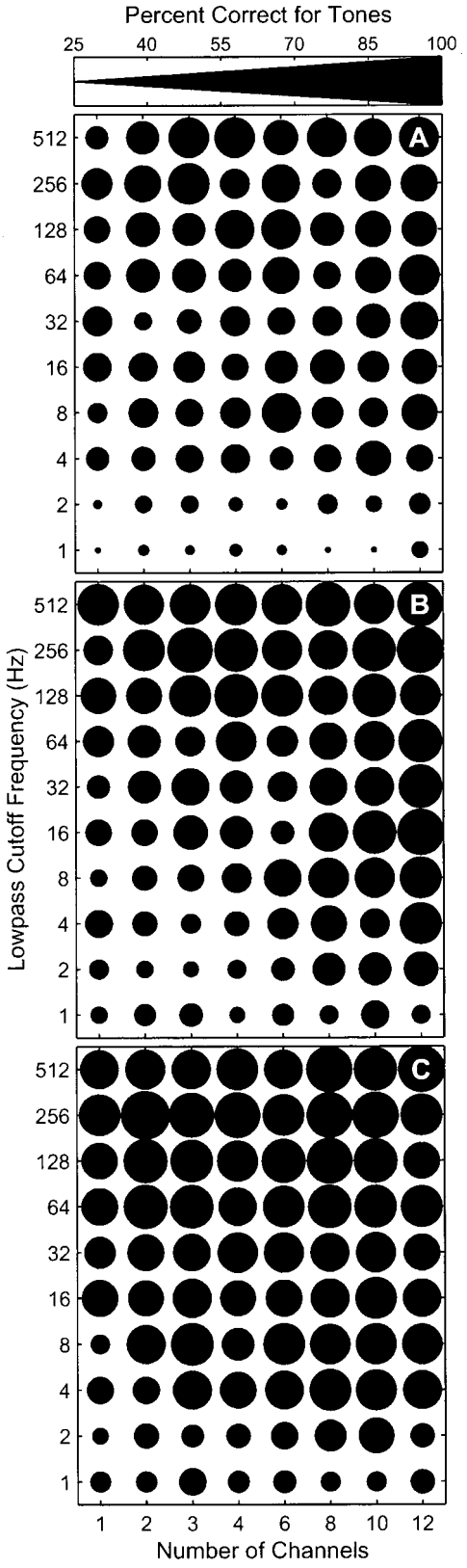


FIG. 6. Representation of the number-of-channels-versus-LPFs matrix of tone recognition scores of three individual subjects. Each panel shows the mean percent correct for tones from one subject. (A) Subject 3. (B) Subject 4. (C) Subject 5. For each panel, the abscissa and the ordinate represent the number of channels and the LPFs, respectively. The percent correct for tones, which ranged between 25% to 100%, is represented by the diameter of the filled circles as indicated by the scale bar at the top.

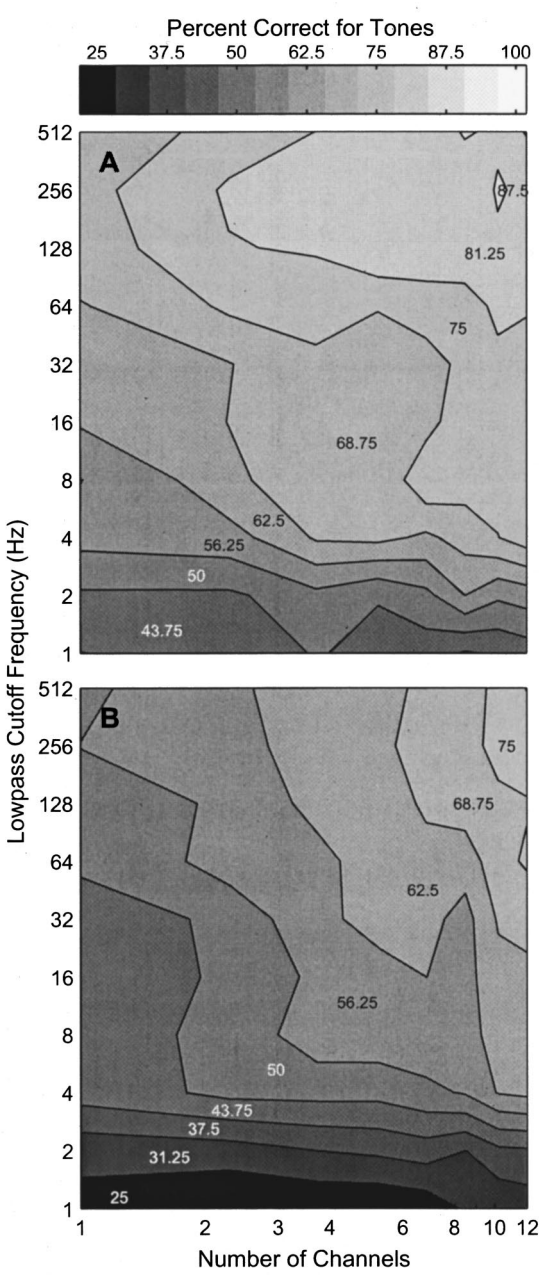


FIG. 7. Representation of the pooled results for the number-of-channels-versus-LPFs matrix of tone recognition scores. The data are plotted in the contour format in which the percent correct is represented by the gray scale as indicated by the scale bar at the top. The abscissa and ordinate are both in logarithmic scales. (A) Data represent the average across all nine subjects who participated in the tone recognition tests using speech materials in which the syllable duration was not equalized. (B) Data represent the average across all four subjects who participated in the tone recognition tests using speech materials that had equal syllable duration. In both (A) and (B), a trade-off between the number of channels and the LPFs is evident by the gradient of the tone recognition scores along the main diagonal line.

LPFs for the tone recognition. Using a small number of channels combined with a high LPF, subjects could achieve a tone recognition performance similar to that achieved using a large number of channels but a low LPF.

The pooled results from the nine subjects who participated in this session (session two) of the experiments are shown in Fig. 7(A). Note the trade-off between the number of channels and the LPFs on the tone recognition performance, similar to the patterns seen for individual subjects in Fig. 6. Two-way ANOVA revealed that both number of channels and LPF affected the tone recognition scores ($p < 0.01$). The pooled data also indicated that the tone recognition performance varied little with number of channels in most cases when the number of channels was greater than six. Similarly, the tone recognition performance was not affected by LPFs when the LPFs were $\geqslant 128$ Hz. High levels of performance of tone recognition could only be achieved with either high LPFs or a large number of channels or both. The low levels of performance with LPFs equal to 1 or 2 Hz were around 40 to 50 percent correct, which was still higher than the chance performance of 25%. This result suggested that the tone information was very robust in the speech materials or that subjects were using some additional cue to help categorize the tokens.

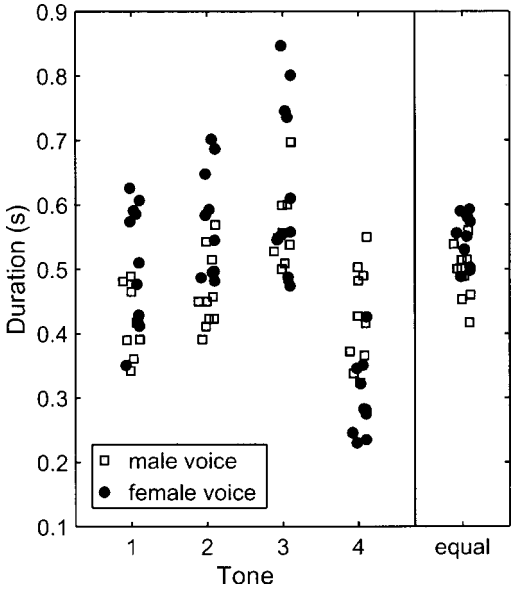


FIG. 8. Syllable durations of tone 1 through 4. Each symbol represents duration of one syllable spoken either by a male voice (open square) or by a female voice (filled circles). The rightmost column, labeled "equal," plots the durations of the syllables that were selected for equal durations for tones 1 through 4.

E. Tone recognition without syllable duration as a potential cue

The durations of Mandarin Chinese syllables vary systematically with the tone patterns, and this could contribute to categorization of the tokens. Figure 8 plots the syllable durations of tone patterns 1 through 4 for speech test materials used in session 2, reported above. There were ten syllables for male voice and ten syllables for female voice for each tone pattern. In general, tone 3 was the longest and tone 4 was the shortest in duration; tone 2 tended to be slightly longer than tone 1. The maximum-likelihood estimator for tone recognition based on the duration alone yielded percent correct of 52.5% for male voice, 62.5% for female voice, and 56.5% when both male and female voices were combined. Note that these levels of performance were only slightly

higher than those observed with one-channel processors where the subjects were forced to use only duration and amplitude cues.

In session 3, using syllables of equal duration for the four tones in the tone recognition test, we found a large decrease in the tone recognition performance for all of the combinations of number of channels and LPFs. Figure 7(B) shows the pooled results from the four subjects who participated in this session of the experiments. Compared to Fig. 7(A), the decreases in performance ranged between 6.0 to 24.9 percentage points with a mean of 15.1 percentage points. Across all tone recognition tests, the decreases in performance were 12.5, 8.5, 19.5, and 19.7 percentage points for tone 1 through 4, respectively. Tone 3 and tone 4 were the longest and shortest in duration (Fig. 8) and removal of the duration cue affected the performance for these two tones the most. These results indicate that the suprasegmental features such as duration of the syllables contributed substantially to the high-level performance of the tone recognition in session 2. Despite the decrease in performance in general when the duration cues were removed, the trade-off between the number of channels and the LPFs was still present (two-way ANOVA, $p<0.01$). For example, to achieve 56.25% correct of tone recognition with one channel, one would need to have LPF$>$256 Hz; however, to achieve the same level of performance with 12 channels, one could use LPF as low as 4 Hz. The other difference between Figs. 7(A) and 7(B) was the ranges of number of channels and LPF on which tone recognition depended. In Fig. 7(A), tone recognition performance increased as number of channels increased from one to six and as LPFs increased from 1 to 128 Hz. In Fig. 7(B), the performance increased as number of channels increased from one to 12 and as LPFs increased from 1 to 256 Hz.

F. Recognition of FM sweeps processed through the cochlear prosthesis simulations

FM sweeps provided an ideal control for duration and temporal envelope, when testing recognition of tone patterns (Fig. 1). It was not surprising that the recognition of the processed FM sweeps was more difficult for the subjects than was recognition of the speech material, because no temporal envelope cues or suprasegmental cues were available to the listeners. Figure 9 shows the mean performance of the four subjects who participated in this session of the experiment. Panels (A) and (B) represent recognition scores for the lower- and higher-pitched FM sweeps processed through the cochlear prosthesis simulation models. Panel (C) represents the pooled data for both lower- and higher-pitched FM sweeps. For the lower-pitched FM sweeps, relatively high performance was achieved at the low-pass cutoff frequency (LPF) of 64 Hz and above [Fig. 9(A)] and relatively little effect of the number of channels were seen. Relatively high performance for the higher-pitched FM sweeps required the LPF of 256 Hz and above [Fig. 9(B)]. In general, little dependency of the performance was found on the number of channels of up to 20. When the number of channels were as high as 30 and 40, however, a remarkable improvement of the recognition performance was noted for the higher-pitched FM sweeps at almost all LPFs compared to the performances for the lower-pitched FM. The trade-off of the LPF and the number of channels for the recognition performance was weak at best for the range of LPFs and number of channels tested.

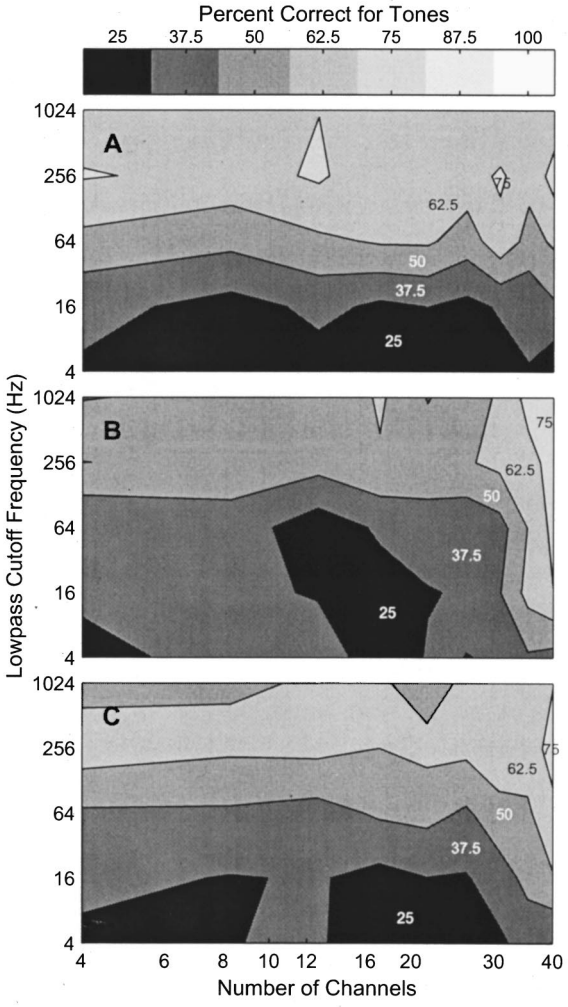


FIG. 9. Mean recognition scores of the four-pattern FM sweeps across all four subjects. The data are plotted in the same format as in one of the panels in Fig. 7 except that the contours are plotted in coarser steps than they are in Fig. 7. Panels (A), (B), and (C) show the percent correct for the lower-, higher-, and both lower- and higher-pitched FM sweeps, respectively. The fundamental frequencies of the FM sweeps are listed in Table I.

IV. DISCUSSION

In normal-hearing subjects listening to acoustic simulations of cochlear prosthesis, four to six independent channels of information are necessary and sufficient to achieve high-level English speech recognition under ideal situations (Shannon *et al.*, 1995; Dorman *et al.*, 1997a; Loizou *et al.*, 1999). Fu and colleagues (1998b) tested speech recognition for Mandarin Chinese in a group of native Mandarin Chinese speakers presented with one- through four-channel acoustic simulations of cochlear prosthesis. The subjects achieved similar consonant, vowel, and sentence recognition to the English counterparts. The tone recognition, however, was found to be independent of the number of channels. The present study extended those observations to a much greater range of channels (i.e., from 1 to 12). Our results were, in

general, consistent with previous observations that the Chinese consonant, vowel, and sentence recognition improved as a function of the number of channels. We also demonstrated the dependence of tone recognition on number of channels, which is discussed below. We found that six channels were necessary for the consonant, vowel, and sentence recognition in quiet to reach optimum performance. Beyond six channels, the vowel recognition improved slightly, but by and large, the consonant-, vowel-, and sentence-recognition performance reached a plateau (Fig. 2).

It should be noted that under more difficult listening conditions, such as in the presence of background noise, the number of channels needed for high-level English speech recognition was found to be much larger, depending on the signal/noise ratio (Dorman *et al.*, 1998; Fu *et al.*, 1998a; Friesen *et al.*, 2001).

There is evidence that cochlear prosthesis can utilize only four to six effective channels (Fishman *et al.*, 1997; Dorman and Loizou, 1998; Fu *et al.*, 1998a; Friesen *et al.*, 2001). In addition, findings from studies using acoustic simulations of cochlear prosthesis in normal-hearing subjects and those from studies in cochlear implant patients are in agreement in that optimum speech recognition in quiet can be achieved with a few channels. This clearly has practical importance when we design an effective, low-cost implant for the developing countries including China (Zeng, 1995; Wilson *et al.*, 1998). We found that high scores of tone recognition with a few channels were achieved using the speech materials in which the syllable duration was not equalized for the four tone patterns. However, syllable duration clearly contributed to the relatively high performance in tone recognition (Fig. 8). Studies using information transmission analysis showed that implant patients were able to utilize duration cues in consonant (van Wieringen and Wouters, 1999) and vowel (Tyler *et al.*, 1992; van Wieringen and Wouters, 1999) recognition when consonants and vowels were presented in isolation. In everyday speech, syllable duration is not a reliable cue for speech recognition. Different speakers might talk at different speeds. The same speaker might change speed depending on many factors, such as the emotional state of the speaker or the context of the sentences. While listeners might adapt to these changes to some extent, variability in talker speed is likely to weaken the duration cues. When we tested the tone recognition using syllables of equal duration for the four tone patterns to eliminate duration as a cue, we found the tone recognition scores dropped an average of about 15 percentage points. Such a decrease in performance was observed within the range of numbers of channels (i.e., 1 to 12) that we tested (Fig. 7).

The best tone recognition scores with 12 channels and a 512-Hz LPF when duration cues were removed were only about 75% correct. In a preliminary report, Wei and colleagues (2001) found that tone recognition of the best performer of a group of four native Chinese speaking patients with cochlear implants was about 80% correct with the number of channels being 7 or 10. Thus, there is room for improvement in tone perception by the cochlear implant users. Current speech coding strategies are not designed to optimize the tone perception. Improvement of pitch perception for implant users is currently an active research goal (e.g., Faulkner *et al.*, 2000; Geurts and Wouters, 2001). That research is applicable to tonal speech and should profit from research done in the tone language speakers.

Although subjective sound quality is important to implant patients, few studies in the field of cochlear implant research have systematically examined the subjective judgments of the sound quality of electric hearing in patients or acoustic hearing of simulations of cochlear implants in normal hearing subjects. Pfingst and colleagues (2001) indicated that the subjective judgments of the quality were only moderately correlated with the subject's speech recognition ability, which suggested that the mechanisms contributing to sound quality and speech recognition ability do not completely overlap. Fishman and colleagues (1997) evaluated the subjective benefit in implant patients of experimental processors that varied in the number of channels. They found that the subjective judgments of the benefit with number of channels up to 10 exceeded that with fewer channels. Little further benefit was seen with 20 channels compared to 10 channels. In the present study we demonstrated that the normal hearing subjects' preferences of the sound quality increased as a function of the simulated number of channels up to 12, the highest number tested (Fig. 3). Caution should be exerted, however, when extrapolating the results from normal-hearing subjects to implant patients. One of the reasons is that we do not know if it is possible to activate a large number of *independent* channels in current cochlear prostheses. Functional channel overlap might negate the potential benefit of a large number of channels. Nevertheless, the present study indicates that even though the speech recognition scores in quiet have reached a plateau at four to six channels, more independent channels can still bring about improvement of the sound quality.

Tone patterns of tone languages are defined by patterns of the fundamental frequency ($F0$) for the voiced part of the word (see Fig. 1). Yet, $F0$ itself does not seem to be essential for tone pattern recognition. Liang (1963) tested the tone recognition of high-passed Mandarin Chinese speech signals. He found that high-level tone recognition was preserved for signals that were high-passed above 2.4 kHz. One might attribute the tone recognition to the residue pitch (Schouten *et al.*, 1962). In another experiment, however, Liang (1963) found that the tone recognition maintained at 60- to 70-percent correct for whispered speech in which neither $F0$ nor its harmonics existed. Whalen and Xu (1992) pointed out that the Mandarin tone information might be carried by the amplitude contour (temporal envelope) of the speech signals. Their subjects maintained high-level recognition of tones when the natural speech was transformed to signal-correlated noise (Schroeder, 1968) in which no $F0$ or formant structure existed but the original amplitude contour was retained. Fu and colleagues (1998b) found that tones were consistently recognized at about the 80%-correct level using acoustic simulations of one- to four-channel CIS-type processing strategies. These results supported the hypothesis that the temporal envelope carries information about the tone patterns. Therefore, as discussed below, under situations of restricted spectral information as in cochlear prostheses, faith-

ful transmission of temporal envelope information would be important for tone recognition.

The present study confirmed the results of Fu *et al.* (1998b) in that high-level recognition of tones could be achieved with a small number of channels and that the more robust temporal information in the envelope (i.e., LPF of 500 Hz versus 50 Hz) was beneficial for tone recognition. Fu and colleagues (1998b) indicated that the tone recognition was not dependent on the number of channels (i.e., one through four). However, our results demonstrated that the tone recognition depended on the number of channels, although the dependence was not as strong as that for vowel or consonant recognition (Fig. 2). In the Fu *et al.* (1998b) study, only a male voice was used for the tone recognition test. In the present study, both male and female voices were used and thus the test became more difficult for the subjects, so that their performance with one or two channels did not reach as high a level as with more channels. When the syllable duration was equalized, the tone recognition became even more challenging and the dependence on the number of channels was more apparent (Fig. 7). Results from cochlear implant users support the conclusion that tone recognition is affected by the number of channels (Wei *et al.*, 2001).

When the spectral information is minimized, the temporal information of the speech signal has been found to contribute to consonant recognition, especially for voicing and manner (Van Tasell *et al.*, 1987; ter Keurs *et al.*, 1992, 1993; Drullman *et al.*, 1994) and for certain aspects of vowel recognition, e.g., short versus long vowels (Rosen, 1992; Drullman *et al.*, 1994). However, little is known about the effects of interaction between spectral cues and temporal cues on speech recognition. For tone recognition, we found that there was a trade-off between temporal and spectral information. The trade-off existed between the number of channels, ranging from one to six or eight, and the LPFs, ranging from 1 to 128 Hz [Fig. 7(A)]. Beyond these, the tone recognition became insensitive to either an increase in number of channels or an increase in the LPFs. The higher LPFs (i.e., 256 and 512 Hz) should permit some or all periodicity information to pass through the acoustic simulations, yet the tone recognition showed no more improvement. This might be due to the already high-level recognition at LPF of 128 Hz leaving little room for improvement. Another interesting observation was that the tone recognition with LPFs of 1 or 2 Hz was 40- to 50 percent correct [Fig. 7(A)], nearly twice as high as the chance performance of 25 percent correct. This suggests that the suprasegmental features of the speech signal (Lehiste, 1970), such as duration of the syllables, probably contribute to the recognition among tone patterns. We specifically addressed this issue by using equal-duration syllables for the tone recognition tests. Under these conditions, we found that the trade-off between the temporal and spectral information for tone recognition extended up to 12 channels and a LPF of 256 Hz [Fig. 7(B)]. Such a trade-off between the number of channels and the LPFs essentially disappeared when the subjects were tested on the recognition of the processed FM sweep patterns (Fig. 9). This is not surprising because there is no temporal envelope or amplitude contour information in those signals. The high LPFs (64 Hz for the lower-pitched FM sweeps and 256 Hz for the higher-pitched FM sweeps) permit the periodicity information to pass through the simulation. In fact, the subjects reported that they could hear the buzz of pulse trains when the LPFs were above the $F0$ of the FM sweeps. When the LPFs were near the $F0$ of FM patterns, the subjects could perform the task by listening to a transition of a buzz sound to a hiss sound (pattern 2) or a transition of a hiss sound to buzz sound (pattern 4). The improved recognition at 30 and 40 channels was probably due to spectral information that was made available to the listeners through the simulations. Since the frequency spacing of the harmonics of the higher-pitched FM sweeps are twice as large as the spacing of the harmonics of the lower-pitched FM sweeps [Fig. 1(B)], simulations using 30 to 40 channels resolve some of the spectral cues for the higher-pitched FM sweeps but much less for the lower-pitched FM sweeps.

V. CONCLUSIONS

With acoustic simulation of a CIS processor, high levels of speech recognition of Mandarin Chinese in quiet could be achieved with four to six channels, although the sound quality improved further with more channels. Good tone recognition was also achieved with a few channels given robust temporal information (e.g., LPFs$>$100 Hz). However, the high-level tone recognition might have been due in part to suprasegmental cues (such as syllable duration) in the speech materials. When we controlled for those suprasegmental cues, the tone-recognition performance dropped markedly. Tone recognition depended on both the number of channels and the LPFs, and such dependency was reflected in a trade-off between the spectral and temporal cues for the tone recognition. When the temporal envelope cue was removed, as in the case of recognition of processed FM patterns, performance drastically deteriorated and in such a situation, spectral information presented in large numbers of channels ($\geqslant$30) and temporal information presented in high LPF ($>$64 Hz) could be useful for tone recognition.

ACKNOWLEDGMENTS

We are grateful to Qianjie Fu and Xiaosong Wang for providing some of the speech testing materials used in this study. John Middlebrooks, San Duanmu, and Kevin Franck provided valuable comments on an earlier version of the manuscript. We also thank Bob Shannon and an anonymous reviewer for their constructive comments. We acknowledge the technical assistance from Jianzhong Lu and Julie Oldfield. The work was supported by NIH/NIDCD Grant Nos. F32 DC00470, R01 DC03808, and T32 DC00011.

Cochlear Corporation (**1999**). Nucleus Technical Reference Manual Z43470 Issue 1, Lane Cove, Australia, Cochlear Limited.

Dorman, M. F., and Loizou, P. C. (**1998**). ''The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels,'' Ear Hear. **19**, 162–166.

Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (**1998**). ''The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels,'' J. Acoust. Soc. Am. **104**, 3583–3596.
Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997a**). ''Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise outputs,'' J. Acoust. Soc. Am. **102**, 2403–2411.
Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997b**). ''Simulating the effect of cochlear-implant electrode insertion depth on speech understanding,'' J. Acoust. Soc. Am. **102**, 2993–2996.
Drullman, R., Festen, J. M., and Plomp, R. (**1994**). ''Effect of temporal envelope smearing on speech reception,'' J. Acoust. Soc. Am. **95**, 1053–1064.
Faulkner, A., Rosen, S., and Smith, C. (**2000**). ''Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants,'' J. Acoust. Soc. Am. **108**, 1877–1887.
Fishman, K. E., Shannon, R. V., and Slattery, W. H. (**1997**). ''Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor,'' J. Speech Lang. Hear. Res. **40**, 1201–1215.
Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, Princeton, NJ).
Friesen, L., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). ''Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants,'' J. Acoust. Soc. Am. **110**, 1150–1163.
Fu, Q. J., Shannon, R. V., and Wang, X. (**1998a**). ''Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,'' J. Acoust. Soc. Am. **104**, 3586–3596.
Fu, Q. J., Zeng, F. G., Shannon, R. V., and Soli, S. D. (**1998b**). ''Importance of tonal envelope cues in Chinese speech recognition,'' J. Acoust. Soc. Am. **104**, 505–510.
Geurts, L., and Wouters, J. (**2001**). ''Coding of the fundamental frequency in continuous interleaved sampling for cochlear implants,'' J. Acoust. Soc. Am. **109**, 713–726.
Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).
Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.
Gu, R. (**1988**). ''Staggered spondaic word test and the competing sentence test,'' Chin. J. Otorhinolaryngol. **23**, 97–99.
Huang, T. S., Wang, N. M., and Liu, S. Y. (**1996**). ''Nucleus 22-channel cochlear mini-system implantations in Mandarin-speaking patients,'' Am. J. Otol. **17**, 46–52.
Keppel, G. (**1991**). *Design and Analysis: A Researcher's Handbook*, 3rd ed. (Prentice-Hall, Englewood Cliffs, NJ).
Liang, Z. A. (**1963**). ''Tonal discrimination of Mandarin Chinese,'' Acta Physiol. Sinica **26**, 85–91.
Lehiste, I. (**1970**). *Suprasegmentals* (MIT Press, Cambridge, MA).
Loizou, P. C., Dorman, M., and Tu, Z. (**1999**). ''On the number of channels needed to understand speech,'' J. Acoust. Soc. Am. **106**, 2097–2103.
Pfingst, B. E., Franck, K. H., Xu, L., Bauer, E. M., and Zwolan, T. A. (**2001**). ''Effects of electrode configuration and place of stimulation on speech perception with cochlear prostheses,'' J. Assoc. Res. Otolaryngol. **2**, 87–103.
Rosen, S. (**1992**). ''Temporal information in speech: Acoustic, auditory and linguistic aspects,'' Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.
Schouten, J. F., Ritsma, R. J., and Cardoz, B. L. (**1962**). ''Pitch of the residue,'' J. Acoust. Soc. Am. **34**, 1418–1424.
Schroeder, M. R. (**1968**). ''Reference signal for signal quality studies,'' J. Acoust. Soc. Am. **44**, 1735–1736.
Seligman, P., and McDermott, H. (**1995**). ''Architecture of the Spectra 22 speech processor,'' Ann. Otol. Rhinol. Laryngol. **104** (Suppl. 166), 139–141.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.
Shannon, R. V., Zeng, F. G., and Wygonski, J. (**1998**). ''Speech recognition with altered spectral distribution of envelope cues,'' J. Acoust. Soc. Am. **104**, 2467–2476.
Skinner, M. W., Clark, G. M., and Whitford, L. A. (**1994**). ''Evaluation of a new Spectral Peak coding strategy for the Nucleus 22 Channel Cochlear Implant System,'' Am. J. Otol. **15** (Suppl. 2), 15–27.
Studebaker, G. A. (**1985**). ''A rationalized arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.
Sun, J. C., Skinner, M. W., Liu, S. Y., Wang, F. N. M., Huang, T. S., and Lin, T. (**1998**). ''Optimization of speech processor fitting strategies for Chinese-speaking cochlear implantees,'' Laryngoscope **108**, 560–568.
ter Keurs, M., Festen, J. M., and Plomp, R. (**1992**). ''Effect of spectral envelope smearing on speech reception. I,'' J. Acoust. Soc. Am. **91**, 2872–2880.
ter Keurs, M., Festen, J. M., and Plomp, R. (**1993**). ''Effect of spectral envelope smearing on speech reception. II,'' J. Acoust. Soc. Am. **93**, 1547–1552.
Thornton, A. R., and Raffin, M. J. (**1978**). ''Speech-discrimination scores modeled as a binomial variable,'' J. Speech Hear. Res. **21**, 507–518.
Tyler, R. S., Preece, J. P., Lansing, C. R., and Gantz, B. J. (**1992**). ''Natural vowel perception by patients with the Ineraid cochlear implant,'' Audiology **31**, 228–239.
Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). ''Speech waveform envelope cues for consonant recognition,'' J. Acoust. Soc. Am. **82**, 1152–1161.
van Wieringen, A., and Wouters, J. (**1999**). ''Natural vowel and consonant recognition by Laura cochlear implantees,'' Ear Hear. **20**, 89–103.
Wei, C. H., Zeng, F. G., Cao, K. L., and Wang, Z. Z. (**2001**). ''Rate discrimination and tone recognition in Chinese cochlear implant patients,'' Assoc. Res. Otolaryngol. Abstr. **24**, 293.
Wei, W. I., Wong, R., Hui, Y., Au, D. K. K., Wong, B. Y. K., Ho, W. K., Tsang, A., Kung, P., and Chung, E. (**2000**). ''Chinese tonal language rehabilitation following cochlear implantation in children,'' Acta Oto-Laryngol. **120**, 218–221.
Whalen, D. H., and Xu, Y. (**1992**). ''Information for Mandarin tones in the amplitude contour and in brief segments,'' Phonetica **49**, 25–47.
Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). ''Better speech recognition with cochlear implants,'' Nature (London) **352**, 236–238.
Wilson, B. S., Rebscher, S., Zeng, F. G., Shannon, R. V., Loeb, G. E., and Lawson, D. T. (**1998**). ''Design for an inexpensive but effective cochlear implant,'' Otolaryngol.-Head Neck Surg. **118**, 235–241.
Xu, S. A., Dowell, R. C., and Clark, G. M. (**1987**). ''Results for Chinese and English in a multichannel cochlear implant patient,'' Ann. Otol. Rhinol. Laryngol. **96** (Suppl. 128), 126–127.
Zeng, F. G. (**1995**). ''Cochlear implants in China,'' Audiology **34**, 61–75.
Zhang, H., Zhao, K. L., and Wang, Z. Z. (**1988**). *MACC: Chinese Minimal Auditory Capability Test* (Beijing Union Hospital, Beijing, P. R. China).

Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners[a)]

Sarah Hargus Ferguson[b)] and Diane Kewley-Port[c)]
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

(Received 4 January 2001; accepted for publication 2 April 2002)

Several studies have demonstrated that when talkers are instructed to speak clearly, the resulting speech is significantly more intelligible than speech produced in ordinary conversation. These speech intelligibility improvements are accompanied by a wide variety of acoustic changes. The current study explored the relationship between acoustic properties of vowels and their identification in clear and conversational speech, for young normal-hearing (YNH) and elderly hearing-impaired (EHI) listeners. Monosyllabic words excised from sentences spoken either clearly or conversationally by a male talker were presented in 12-talker babble for vowel identification. While vowel intelligibility was significantly higher in clear speech than in conversational speech for the YNH listeners, no clear speech advantage was found for the EHI group. Regression analyses were used to assess the relative importance of spectral target, dynamic formant movement, and duration information for perception of individual vowels. For both listener groups, all three types of information emerged as primary cues to vowel identity. However, the relative importance of the three cues for individual vowels differed greatly for the YNH and EHI listeners. This suggests that hearing loss alters the way acoustic cues are used for identifying vowels. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1482078]

PACS numbers: 43.71.Es, 43.70.Fq, 43.71.Ky, 43.66.Ts [CWT]

I. INTRODUCTION

Many studies and most models of vowel perception have focused on the frequencies of the first two formants (F1 and F2) measured at the vowel's nominal steady state. While it is clear that this spectral target is an important cue to vowel identity, other studies have demonstrated that vowel duration and dynamic formant movement also influence categorization of vowels. For example, in a discriminant analysis of a large multitalker database, Hillenbrand *et al.* (1995) found that the F1–F2 target correctly classified vowels only 68% of the time. This result is unimpressive when compared with the performance of human listeners, who identified the same vowels with about 95% accuracy in quiet. When Hillenbrand *et al.* measured F1 and F2 at two time locations rather than only at the steady state (providing the classifier with dynamic formant information), accuracy improved by 20 percentage points. Adding duration improved performance even further, although not as much as adding dynamic information.

The Hillenbrand *et al.* (1995) results corroborate data from other studies demonstrating that in some cases, formant dynamics and vowel duration may be as important as the spectral target for cueing vowel identity. For example, several studies have used silent-center vowel stimuli, in which the steady-state portion of vowels was replaced with silence. In general, these studies have shown equivalent performance for stimuli containing only the onset and offset transitions and for stimuli containing only the vowel center (Strange, 1989). Thus, Hillenbrand *et al.* found the spectral target insufficient to account for the high intelligibility of vowels, and silent-center studies suggest the spectral target is not necessary for vowel identification.

In light of these results, recent research has sought to determine the relative importance of three types of acoustic information to vowel identity: (1) spectral target information, (2) dynamic formant movement, and (3) duration. For example, Neel (1998) assessed this relationship for normal-hearing and hearing-impaired listeners. Specifically, Neel examined vowel categorization for resynthesized vowel stimuli with and without duration and formant dynamic cues. Neel found that both normal-hearing and hearing-impaired listeners identified full-cue stimuli (those having dynamic formant movement and appropriate duration) substantially better than they did stimuli containing only spectral target cues. The benefit gained from dynamic and duration information was similar for the two groups. The duration result was consistent with Summers and Leek (1992), who found that normal-hearing and hearing-impaired subjects made similar use of duration in identifying synthetic vowels as /i/ or /ɪ/. In contrast, the hearing-impaired listeners in Summers and Leek (1992) relied less on F2 than normal-hearing listeners did. The authors suggested that this adjustment in perceptual weighting may be related to poor frequency resolution in the hearing-impaired subjects.

The current study continued this exploration of the relative importance of acoustic cues to vowel identification for normal-hearing and hearing-impaired listeners. However, in contrast with previous studies, the current study employed natural speech elicited in such a way as to achieve a natural distribution of intelligibility among tokens of each vowel. One way to vary intelligibility is for the talker to vary his or

[a)]Portions of these data were presented at the 138th meeting of the Acoustical Society of America [J. Acoust. Soc. Am. **106**, 2272 (1999)].
[b)]Electronic mail: safergus@indiana.edu
[c)]Electronic mail: kewley@indiana.edu

her speaking style. Several studies have shown that when a talker is instructed to speak as though he or she were talking to a hearing-impaired person, the resulting ''clear'' speech is significantly more intelligible than his or her usual conversational speech (Gagné *et al.*, 1995; Helfer, 1998; Payton *et al.*, 1994; Picheny *et al.*, 1985; Schum, 1996; Uchanski *et al.*, 1996). The reported intelligibility advantage for clear speech has ranged from 11 to 34 percentage points, but in general has been independent of presentation level, presence vs absence of frequency-dependent amplification (Picheny *et al.*, 1985), degradation by noise or reverberation (Payton *et al.*, 1994), or whether listeners were normal-hearing subjects listening in broadband noise or hearing-impaired subjects listening in quiet (Payton *et al.*, 1994; Uchanski *et al.*, 1996).

Picheny *et al.* (1986) reported several acoustic differences between clear and conversational speech. Most notably, clear speech is substantially slower than conversational speech. The slowness of clear speech results not only from longer and more numerous pauses, but also from lengthening of phonetic segments. For example, in the Picheny *et al.* (1986) corpus, short vowels were 29% longer in clear speech than in conversational speech, while unvoiced fricatives were 91% longer in clear speech (Uchanski *et al.*, 1996). Two studies have explored the extent to which longer pause and segment duration accounts for the superior intelligibility of clear speech. Picheny *et al.* (1989) and Uchanski *et al.* (1996) used different signal processing techniques to give conversational speech the durational characteristics of clear speech. In neither study did this processing actually improve intelligibility. Nejime and Moore (1998) also found no intelligibility benefit from artificially slowed speech. Thus, acoustic characteristics other than segment and pause duration may underlie the clear speech advantage.

Relevant to the current study, vowels have been shown to occupy a larger F1×F2 formant frequency space in clear speech than in conversational speech (Picheny *et al.*, 1986). Johnson *et al.* (1993) also found a larger F1×F2 space in so-called ''hyperarticulated'' speech than in citation-style speech. Similarly, Moon and Lindblom (1994) found considerably less vowel undershoot (deviation from expected formant targets) in clear speech than in citation-style speech, in addition to longer duration. The previous data thus indicate two ways in which vowels in clear speech differ from those in conversational speech. First, they are longer in duration; second, the spectral targets occur at more extreme locations in the vowel space. In addition, while no prior study has investigated how speaking style affects dynamic formant movement, pilot data from our laboratory suggest that vowels in clear speech have more extensive formant movement than vowels in conversational speech. Taken together, these data suggest that speaking style manipulations may serve as a window on the relative importance of spectral target, formant dynamic, and duration cues to vowel identification.

Previous investigations of clear speech have primarily examined sentence intelligibility in background noise and have only partially examined acoustic differences between clear and conversational speech. The present experiments extend this line of research to vowel intelligibility and include detailed acoustic analyses of vowels in the two speaking styles. The four experiments described here explored four hypotheses.

(1) Vowels in clear speech would be significantly more intelligible than vowels in conversational speech, for both normal-hearing (experiment 1) and hearing-impaired subjects (experiment 3).
(2) Vowels in clear versus conversational speech would differ significantly in their acoustic characteristics (experiment 2).
(3) Despite significant perceptual and acoustic differences between clear and conversational speech, intelligibility of individual vowel tokens would yield variance sufficient for regression analyses of the relationship between acoustic cues and vowel intelligibility for normal-hearing (experiment 2) and hearing-impaired (experiment 4) listeners.
(4) The relationship between acoustic properties and intelligibility of vowels for hearing-impaired listeners would differ from that observed for normal-hearing listeners (experiment 4).

II. CLEAR AND CONVERSATIONAL SPEECH MATERIALS

Test words for all experiments consisted of 10 vowels (/i, ɪ, e, ɛ, æ, ɑ, ʌ, o, ʊ u/) in /b*V*d/ context. Each test word was centered in one of 12 neutral carrier sentences for elicitation. A list of 140 sentences was generated that included five instances of each vowel in /b*V*d/ context (a different sentence for each occurrence) and one instance in each of nine other CVC contexts. The CVC materials were used in a separate project.

Recordings were made of six randomizations of this list. The talker, a 59-year-old male audiologist, was chosen for his dialect (Central Indiana) and for his extensive experience communicating with hearing-impaired people. In the first recording session, he was instructed to read the first three randomizations of the sentence list using his normal, conversational speaking style. At a second recording session on the following day, he was instructed to read the second three randomizations, speaking as though he were talking to a hearing-impaired person, taking care to enunciate each word carefully. The talker rehearsed each style prior to recording and the experimenter monitored his productions for errors during recording. Recordings were made in a single-wall, sound-attenuating booth using a Shure SM-10 headset microphone placed approximately 1 inch from the talker's lips. Microphone output was routed to a preamplifier (Shure M267) and then to a digital audio tape (DAT) recorder (Panasonic SV 3800).

To generate test stimuli, the sentences were low-pass filtered (8500 Hz) and digitized (16-bit A/D, 22 050 Hz sampling rate). Test words were then excised from the sentences via computer editing. For each speaking style, 10 of the 15 tokens recorded for each vowel were chosen as stimuli. The stimuli were arbitrarily selected from the second and third randomizations; any tokens judged to be contaminated by noise were replaced with a token from the first randomiza-

tion. To eliminate amplitude differences among vowels and between the two speaking styles, the average RMS amplitude was calculated across all test items, and then all items were scaled to this amplitude using Matlab. Informal testing in quiet with normal-hearing listeners demonstrated all test items to be clearly identifiable as the intended vowel.

For the perceptual experiments, test stimuli were delivered in a background of 12-talker babble. A 30-s sample of babble was low-pass filtered and digitized from the noise channel of a recording of the Speech Perception in Noise Test (Kalikow *et al.*, 1977). Digitization parameters were identical to those used for the sentences. On each test trial, a segment of babble was selected from a random location within this 30-s sample. The duration of this segment exceeded that of the test item by 1000 ms such that the test word was centered temporally in the babble.

Along with the vowel materials described above, two additional lists were recorded, each consisting of 14 sentences selected from the Central Institute for the Deaf (CID) Everyday Sentences test (Davis and Silverman, 1978). The first list was recorded in a conversational speaking style, immediately after the vowel materials were recorded in that style. The second list was recorded in a clear speaking style. The CID sentences were presented to normal-hearing pilot subjects at a comfortable presentation level in 12-talker babble (speech-to-babble ratio = 0 dB). The average percent words correct was 80% for clear sentences and 46% for conversational sentences, a difference that was significant in a paired *t*-test ($p < 0.01$). The clear speech benefit on this pilot test (34 percentage points) was within the range reported by previous studies on clear speech. This confirmed that for the current talker, sentences produced under instructions to speak clearly were significantly more intelligible than sentences produced conversationally.

III. EXPERIMENT 1: VOWEL INTELLIGIBILITY FOR YOUNG, NORMAL-HEARING (YNH) LISTENERS

A. Method

1. Subjects

Nine young adults (20–23 years) with normal hearing (thresholds ≤ 20 dB HL at half-octave intervals from 250–8000 Hz) participated in this experiment. All YNH listeners were native speakers of General American English with no history of speech or language disorders.

2. Procedures

Listeners performed all testing in a single-wall sound treated room, seated in front of a computer monitor and keyboard. On each trial, a test word and babble segment were played out from separate channels of a 16-bit D/A converter [Tucker–Davis Technologies (TDT) DA1]. The speech was attenuated (TDT PA4) to achieve the desired speech-to-babble (S/B) ratio. The speech and babble were then mixed (TDT SM3) and routed to a programmable filter (TDT PF1) that low-pass filtered (8500 Hz) and attenuated the combined signal to the desired overall speech level. The speech and babble were then delivered to the subject's right ear via ER-3A insert earphones. The subject identified the vowel of the test word by typing the number of the response category corresponding to that vowel. Response alternatives were displayed on the monitor as keywords[1]: (1) feet, bead; (2) sit, rib; (3) tape, raid; (4) pet, head, said; (5) back, mad; (6) pot, sod; (7) cup, bug; (8) rode, coat; (9) book, would; and (10) rude, news, food.

YNH subjects were tested in four sessions, each lasting approximately 90 minutes. In the first session, listeners' hearing was screened and they were familiarized with the test procedures prior to beginning the experimental conditions. For familiarization, a 40-trial block of clear vowel tokens was created using materials from a previous project. YNH listeners heard this block two times: in quiet with feedback, and then in babble ($S/B = -5$ dB) without feedback. All YNH listeners identified vowels with 90% or greater accuracy in quiet. For the experimental conditions, test stimuli were arranged into four blocks of 100 items: clear /bVd/, conversational /bVd/, clear mixed CVCs, and conversational mixed CVCs (only /bVd/ data are described in the current report). Listeners heard each block four times, for 16 total test blocks. Stimuli were randomized within each block and each subject received the 16 test blocks in random order. Test words were presented at an overall level of 70 dB SPL with an S/B ratio of −10 dB. In pilot tests, this S/B ratio was shown to prevent ceiling effects for the most intelligible vowels. Feedback was not given during the test blocks.

TABLE I. Mean percent correct intelligibility for vowels in clear (CL) and conversational (CNV) speech. Standard deviations are shown in parentheses. The speaking style effect was significant ($p < 0.05$) for YNH listeners but not EHI listeners.

Subjects	Speech, dB SPL	S/N ratio, dB	Style	% corr.
YNH	70	−10	CNV	71.6 (5.64)
	70	−10	CL	86.6 (3.60)
EHI	70	−3	CNV	73.8 (14.31)
	70	−3	CL	73.1 (13.11)

B. Results and discussion

Intelligibility data for experiment 1 are shown in the top half of Tables I and II. Table I displays percent correct scores averaged across listeners and vowels; test parameters also appear in this table. Table II shows data for individual vowels. Vowels in clear speech (CL) were more intelligible than vowels in conversational speech (CNV), by an average of 15 percentage points. Percent correct scores were converted to rationalized arcsine units (RAUs) prior to statistical analysis (Studebaker, 1985). The effects of speaking style and vowel were analyzed in a two-way repeated-measures analysis of variance (ANOVA) using Statistica (1999).

The main effect of speaking style was significant [$F(1,8) = 96.47$, $p < 0.0001$]. The size of the clear speech benefit observed here (15 percentage points) is comparable to that found in previous studies with normal-hearing listen-

TABLE II. Mean percent correct intelligibility for individual vowels in clear (CL) and conversational (CNV) speech (see Table I for test conditions). Standard deviations are shown in parentheses. Significant speaking style effects ($p<0.05$) are denoted by boldface.

Group	Style	/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ʌ/	/o/	/ʊ/	/u/
YNH	CNV	90.8 (11.79)	93.1 (5.12)	**78.9** (8.67)	**47.2** (18.73)	**31.4** (20.92)	95.0 (3.75)	**71.7** (13.46)	**71.7** (5.86)	**68.6** (12.38)	67.8 (18.18)
	CL	91.7 (4.68)	90.3 (2.92)	**91.4** (3.97)	**76.4** (9.53)	**85.3** (10.34)	97.8 (2.64)	**93.6** (4.70)	**83.9** (7.61)	**81.1** (7.92)	74.7 (13.60)
EHI	CNV	**68.1** (25.12)	58.9 (26.67)	**78.1** (19.60)	65.3 (20.21)	78.3 (20.04)	91.1 (11.67)	**60.3** (23.57)	88.6 (7.92)	80.6 (14.46)	**69.2** (22.53)
	CL	**41.7** (27.89)	53.3 (26.93)	**66.9** (22.07)	56.1 (21.87)	80.0 (11.32)	86.4 (19.33)	**88.9** (7.51)	86.9 (5.70)	87.5 (9.27)	**83.1** (12.36)

ers. For example, Gagné *et al.* (1995) reported an average clear speech improvement of 13.5 percentage points for natural sentences presented at a 0 dB signal-to-noise (S/N) ratio. Similarly, in Uchanski *et al.* (1996) the clear speech advantage averaged 16 percentage points for nonsense sentences with a S/N ratio of -4.5 dB. The clear speech benefit reported by Payton *et al.* (1994) ranged from 15 to 34 percentage points across a variety of listening conditions, some of which included reverberation and/or combined noises (speech-shaped noise plus white noise). Note that all of these earlier studies used sentence materials; no previous study has assessed vowel identification in clear and conversational speech for normal-hearing listeners.

The main effect of vowel and the interaction between speaking style and vowel also were significant [$F(9,72)=28.19$, $p<0.0001$; $F(9,72)=12.14$, $p<0.0001$]. These effects are shown in Fig. 1. Planned comparisons for speaking style effects by vowel were significant ($p<0.01$) for /e/, /ɛ/, /æ/, /ʌ/, /o/, /ʊ/. For /i/, /ɪ/, and /ɑ/, the lack of a clear speech benefit resulted from relatively high performance for conversational tokens. Post-hoc tests (Tukey HSD) revealed that in conversational speech, these three vowels were significantly more intelligible than the other seven. In the case of /i/ and /ɑ/, this result was not surprising because these vowels are known to be robust under adverse listening conditions.

IV. EXPERIMENT 2: ACOUSTIC AND CORRELATIONAL ANALYSIS

The purpose of experiment 2 was to determine the acoustic characteristics underlying the speaking style effect demonstrated in experiment 1. In brief, the vowel stimuli were analyzed acoustically. Within each vowel, clear and conversational tokens were compared on steady state, dynamic, and durational acoustic characteristics. Regression analyses then were used to explore the relationship between acoustic characteristics and vowel intelligibility.

A. Acoustic analysis

1. Method

For the LPC analysis, test stimuli were downsampled from 22 050 Hz to 11 025 Hz. Formant tracking was accomplished with modified Colea Matlab code (Loizou, 2000) us-

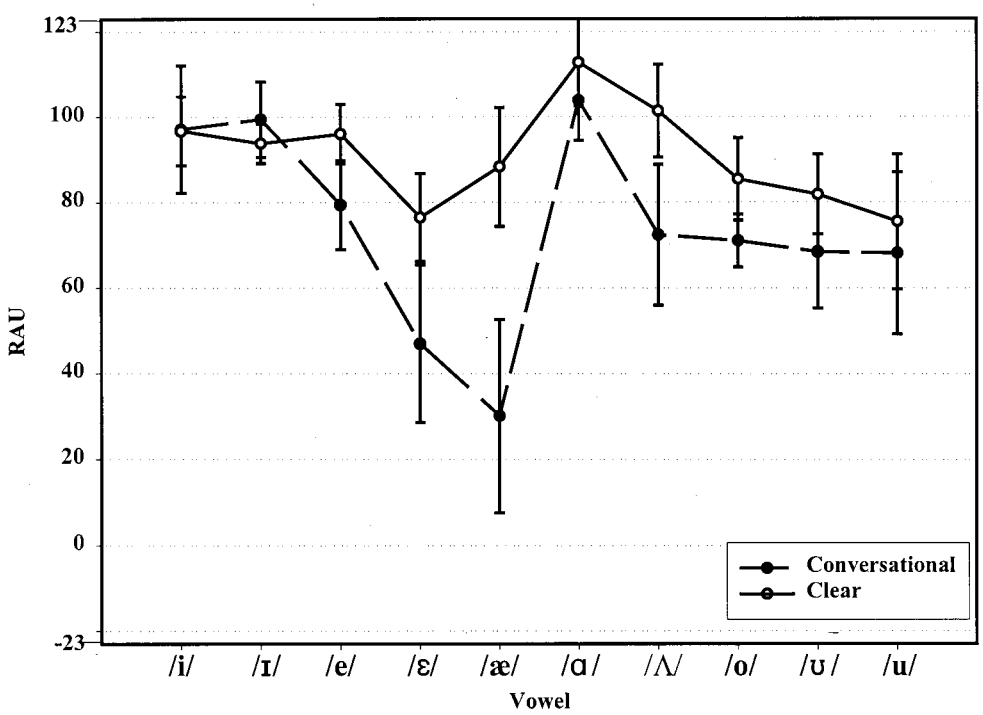


FIG. 1. Vowel identification performance in RAUs for YNH listeners. Error bars are ±1 standard deviations from mean for nine listeners.

TABLE III. Mean absolute acoustic data for clear and conversational tokens of individual vowels. Standard deviations are shown in parentheses. Significant speaking style differences ($p<0.05$) for each metric and vowel are denoted by boldface.

Metric	Style	/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ʌ/	/o/	/ʊ/	/u/
F1 (barks)	CNV	**3.25** (0.24)	4.27 (0.24)	**4.51** (0.11)	5.18 (0.21)	**5.84** (0.15)	**6.73** (0.22)	**5.54** (0.20)	**4.86** (0.20)	**4.33** (0.19)	**3.70** (0.23)
	CL	**3.52** (0.19)	4.21 (0.21)	**5.08** (0.20)	5.25 (0.18)	**6.35** (0.21)	**7.38** (0.12)	**6.24** (0.22)	**5.44** (0.20)	**4.67** (0.27)	**4.26** (0.53)
F2 (barks)	CNV	**13.96** (0.15)	**12.97** (0.26)	**13.29** (0.14)	**12.48** (0.23)	**12.64** (0.14)	9.41 (0.16)	**10.43** (0.11)	8.83 (0.33)	**8.90** (0.30)	9.98 (0.35)
	CL	**14.50** (0.22)	**13.79** (0.30)	**13.68** (0.20)	**13.20** (0.28)	**12.96** (0.15)	10.09 (1.68)	**10.18** (0.23)	8.69 (0.16)	**8.40** (0.26)	9.97 (0.67)
Vector length (barks)	CNV	**0.36** (0.25)	**0.63** (0.50)	**1.33** (0.27)	**0.24** (0.18)	0.27 (0.09)	1.22 (0.15)	1.48 (0.34)	**1.29** (0.47)	**2.07** (0.35)	**0.52** (0.25)
	CL	**0.92** (0.32)	**1.63** (0.38)	**2.43** (0.44)	**0.92** (0.25)	0.29 (0.19)	1.43 (1.10)	1.67 (0.24)	**2.33** (0.33)	**2.50** (0.49)	**2.22** (1.11)
Duration (ms)	CNV	**146.5** (21.72)	**129.4** (43.97)	**203.6** (47.67)	**122.0** (19.08)	**217.2** (23.65)	**255.2** (34.33)	**136.1** (29.66)	**228.7** (43.69)	**200.1** (20.64)	**209.3** (32.18)
	CL	**416.6** (26.21)	**300.4** (33.38)	**427.2** (53.92)	**325.2** (29.23)	**446.1** (32.97)	**468.1** (31.93)	**344.2** (80.72)	**455.5** (28.32)	**340.7** (43.01)	**373.2** (27.07)

ing 12 LPC coefficients, a 20 ms Hamming window, and a 10 ms frame rate. The few tracking errors were corrected by hand editing. Vowel duration measures were made from the wave form. Values of the first two formants (F1 and F2) were extracted from the formant tracks at three temporal locations: 20% of the vowel duration, the vowel steady state (defined as 30 ms after the 20% location), and 80% of the vowel duration.

There were two classes of acoustic metrics: absolute and relative. Absolute metrics were used for assessing acoustic differences between the two speaking styles. They included F1 at steady state (F1SS), F2 at steady state (F2SS), vowel duration (DUR), and vector length (VL). VL, the length of a vector in F1×F2 space connecting the 20% and 80% values of the formants, was calculated in barks (Traunmüller, 1990) as the Euclidean distance. VL indicated magnitude of formant movement over the vowel nucleus following work by Hillenbrand *et al.* (1995). Relative metrics were used for the regression analyses; these are described below.

2. Results and discussion

A summary of the acoustic data is shown in Table III. The values in each cell represent the mean across the 10 tokens in a given vowel/speaking style combination. To compare clear and conversational tokens of each vowel, a multivariate analysis of variance (MANOVA) was performed for each vowel. The four dependent variables were F1SS, F2SS, VL, and DUR. Speaking style was the independent variable. The effect of speaking style was significant for all vowels [$F(4,15)>16$, $p<0.0001$]. Planned comparisons were performed to assess the effect of speaking style for each individual metric and vowel. Most contrasts were significant; these are indicated by boldface in Table III. Figure 2 also shows steady-state formant values for vowels in clear and conversational speech.

Some general trends are apparent in Table III and Fig. 2. First, for the eight vowels (all except /ɪ/ and /ɛ/) where F1 differed significantly between the two speaking styles, F1 was higher in clear speech than in conversational speech. This effect is consistent with recent data demonstrating that F1 rises when talkers increase vocal effort (Liénard and DiBenedetto, 1999). Talkers might be expected to increase their effort when speaking clearly, and previous data suggest that this in fact occurs. Picheny *et al.* (1986) reported that clear speech was 5 to 8 dB more intense than conversational speech. Increased intensity for the clear condition was observed informally in the current study; during recording and digitization, input levels to the DAT and to the computer needed to be decreased to avoid clipping on clear productions.

In contrast with F1, which was generally higher in clear speech, the direction of change for F2 depended on the vowel. For all five of the front vowels (/i/, /ɪ/, /e/, /ɛ/, /æ/), F2 was significantly higher in clear speech than in conversational speech. For the back vowels (/ɑ/, /ʌ/, /o/, /ʊ/, /u/), F2 was, in general, lower in clear speech than in conversational speech, although the speaking style effect was significant only for /ʌ/ and /ʊ/. Overall, then, vowels in clear speech occupied a larger F2 space than vowels in conversational speech. Previous studies suggest that larger vowel spaces are

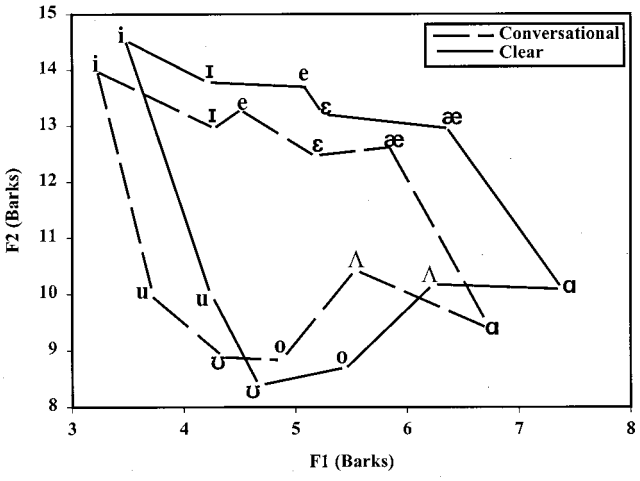


FIG. 2. Average steady-state F1 and F2 frequencies for vowels in clear and conversational speech.

TABLE IV. Summary of regression analysis for YNH subjects. DM=dynamic metric, F1dis=F1 distance, F2dis=F2 distance, TDUR=transformed duration.

	/e/	/ɛ/	/æ/	/ʌ/	/o/	/ʊ/	/u/
Predictor 1	TDUR	DM	TDUR	F1dis	F1dis	F1dis	F1dis
R^2	0.25	0.59	0.73	0.40	0.34	0.22	0.25
Predictor 2	F2dis	TDUR	F1dis	TDUR	F2dis		DM
R^2 change	0.16	0.05	0.03	0.06	0.06		0.05
Predictor 3		F2dis			TDUR		
R^2 change		0.05			0.12		
Predictor 4		F1dis					
R^2 change		0.03					
Multiple R	0.64	0.85	0.87	0.68	0.72	0.47	0.55
p	<0.015	<0.001	<0.001	<0.007	<0.01	<0.04	<0.06

associated with increased intelligibility (e.g., Bradlow *et al.*, 1996).

The vector length data also seem to reveal strategic enhancement of vowel acoustic properties. Vector length was significantly greater in clear speech than in conversational speech for most vowels, but not /æ/, /ɑ/, or /ʌ/. In Fig. 2, the part of the vowel space containing these three vowels appears less crowded than the rest of the vowel space. This suggests that for the relatively more crowded regions of his vowel space (one containing /i/, /ɪ/, /e/, and /ɛ/; the other including /o/, /ʊ/, and /u/) the talker made his vowels more dynamic in clear speech. In the region where his F1×F2 space is less crowded (e.g., for /æ/, /ɑ/, and /ʌ/), he used the same formant movement strategy in both speaking styles.

Finally, the most robust result for these acoustic measures involved vowel duration. All vowels were significantly longer in clear speech than in conversational speech, with vowels in clear speech being 2.1 times longer, on average, than vowels in conversational speech.

In summary, clear and conversational tokens of vowels recorded from a single talker differed significantly on a number of acoustic measures. Of the 10 vowels, all showed significant speaking style effects for duration plus at least one other metric. This result indicates that when given instructions to speak as though talking to a hard-of-hearing person, the talker made specific changes in his vowel production. The concomitant changes in steady-state and dynamic formant frequency data are consistent with previous literature showing that clear speech is not simply a slower version of conversational speech.

B. Regression analysis

1. Method

Relative acoustic metrics were computed within each vowel category in reference to the most intelligible or "best" token (BT), observed across both clear and conversational tokens. For vowels where more than one token shared the highest score, the token with more extreme steady state formant values was designated the BT. Relative metrics were then calculated for the other 19 tokens in that category. For steady-state formant measures, the relative metrics were F1 distance (F1 dis) and F2 distance (F2dis). These signed distances in barks were calculated for each vowel category by subtracting the steady-state value for the BT from the steady-state value for the token in question. Thus, while the transformation was unique for each vowel, in every case F1dis and F2dis were simple linear transformations of absolute F1 and F2 values.

A third relative metric, indicating the degree to which a token resembled the BT in terms of dynamic formant movement, was less straightforward. Following Kewley-Port *et al.* (1996), this dynamic metric (DM) comprised a scalar product of the 20%–80% vectors for the token and the BT. Spectral location was normalized by shifting both vectors to the origin before calculating the scalar product. The DM was then scaled so that its value for the BT was 1.0; values for tokens very similar to the BT in direction and extent approached 1.0. DM values near zero indicated token vectors that were highly dissimilar in direction and/or extent to the BT vector.

As reported above, vowels were twice as long, on average, in clear speech than in conversational speech. For most vowels, this large effect involved a bimodal distribution of duration across the 20 tokens, which violates the assumptions of linear regression. To address this problem and still examine the relationship between duration and intelligibility, duration for the clear tokens was transformed by subtracting 102.45 ms from the original raw duration (the mean difference between the two speaking styles was 204.9 ms). To ensure that the significant speaking style effect observed in the absolute acoustic data was not due solely to extreme duration differences, the MANOVAs were repeated using this transformed duration (TDUR) data in place of the original duration. The overall speaking style effect across the four metrics (F1SS, F2SS, VL, and TDUR) remained significant ($p<0.003$) for all vowels, as did TDUR ($p<0.03$). Absolute TDUR values, not referenced to the BT, were used in the regression analysis.

Vowels were selected for analysis based on three criteria designed to assure sufficient variability in intelligibility for the regression: (1) significant intelligibility difference between the two speaking styles, (2) median token score less than 90% correct, and (3) token score range greater than 40 percentage points. The seven vowels that met two of these three criteria were submitted to a forward stepwise multiple regression analysis: /e/, /ɛ/, /æ/, /ʌ/, /o/, /ʊ/, and /u/. The dependent variable was the token intelligibility score in

RAUs. Three predictors were the relative metrics F1dis, F2dis, and the DM; the fourth was TDUR.

2. Results and discussion

Regression results are summarized in Table IV. For each vowel, the top four rows of the table show the predictor(s) present in the final regression equation of the forward stepwise analysis and the variance, R^2, accounted for by each. Predictors are numbered by the order in which they entered the equation. The bottom two rows of Table IV give the multiple R values of the regression equations and their significance levels.

F1dis was the most important predictor of token intelligibility for four of the seven vowels analyzed, specifically the back vowels /ʌ/, /o/, /ʊ/, and /u/. One of these vowels (/o/) is shown in Fig. 3. In this figure we see that clear tokens (which were more intelligible than conversational tokens by 12 percentage points) had generally higher F1 values than the BT (indicated by a diamond), while conversational tokens tended to have lower F1 values than the BT. For F2, clear tokens were, on average, closer to the BT than conversational tokens. F2dis accounted for some of the variance for /o/ but failed to enter the regression equation for the other three back vowels.

While the back vowels as a group were consistent in having F1dis as the primary predictor of intelligibility, very different patterns of cue use were observed for the three front vowels analyzed (/e/, /ɛ/, and /æ/). For /ɛ/, the strongest predictor of intelligibility was the DM. As seen in Fig. 4, the clear tokens (which were 29 percentage points more intelligible than conversational tokens) resembled the BT in both direction and extent of formant movement, while the conversational tokens had very different formant movement patterns from the BT. Incidentally, a similar pattern was observed for the back vowel /u/, for which the DM accounted for about 5% of additional variance in intelligibility (F1dis accounted for 25%).

In contrast with /ɛ/, the primary predictor of intelligibility for /e/ and /æ/ was TDUR, which accounted for 25% of the variance in intelligibility for /e/ and 73% of the variance for /æ/. That duration would emerge as a primary cue for any vowel was unexpected, because duration is traditionally considered a secondary cue to vowel identity in American English. TDUR was a secondary predictor of intelligibility for /ɛ/, /ʌ/, and /o/, accounting for 5% to 12% of additional variance in each case.

In summary, the regression analysis suggests that for this talker's back vowels, steady-state F1 information was the primary cue that young, normal-hearing subjects used to establish vowel identity. For his front vowels, duration and dynamic formant movement held primary importance, with specific patterns of cue use varying among the three vowels analyzed. Experiments 3 and 4 explore the extent to which the patterns of acoustic cue use observed for this talker were altered by hearing loss.

V. EXPERIMENT 3: VOWEL INTELLIGIBILITY FOR ELDERLY, HEARING-IMPAIRED (EHI) LISTENERS

A. Method

1. Subjects

Nine EHI listeners were recruited from the participants of a large-scale, longitudinal study on hearing aid outcome measures. Subjects in the longitudinal study were 60 to 89 years old and had bilaterally symmetrical (interaural difference $\leqslant$30 dB for 250–4000 Hz) sensorineural hearing loss of flat or gently sloping configuration. Subjects for the present experiment met the additional criterion of mild-to-moderate sloping loss, with threshold limits of 35 dB HL at 500 Hz, 40 dB HL at 1000 Hz, 50 dB HL at 2000 Hz, and 65 dB HL at 4000 Hz. Audiometric thresholds for test ears, along with subject ages, are shown in Table V. Subjects were native speakers of American English, had no noticeable regional accent, and reported no history of speech or language disorders. All scored in the 35th percentile or higher for their age group on the Wechsler Adult Intelligence Scale—Revised (Wechsler, 1983). The ear that best fit our audiometric criteria was chosen as the test ear.

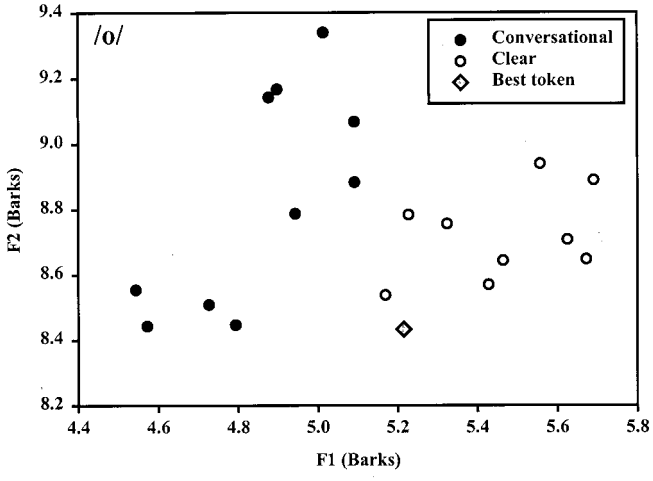


FIG. 3. Steady-state F1 and F2 frequencies for individual tokens of /o/. "Best token" refers to the most intelligible token of /o/ based on YNH intelligibility data.

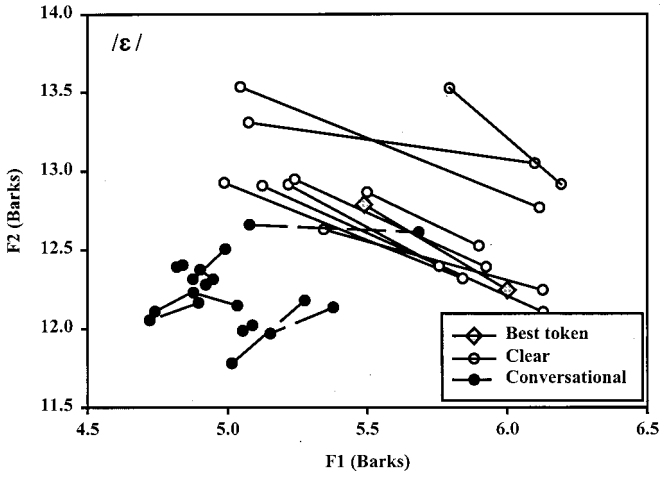


FIG. 4. Dynamic plot for /ɛ/. Each vector represents the distance, for a single token of /ɛ/, between F1 and F2 values at 20% and 80% of the vowel duration. "Best token" refers to the most intelligible token of /ɛ/ based on YNH intelligibility data.

TABLE V. Age and test-ear audiometric thresholds (in dB HL) for EHI subjects.

		Frequency								
Listener	Age	250	500	1000	1500[a]	2000	3000	4000	6000	8000
S1	71	15	10	5	25	45	65	65	75	70
S2	81	10	10	25		35	40	65	55	85
S3	76	10	15	30		45	50	55	70	75
S4	76	5	10	15	25	45	60	60	65	65
S5	62	20	20	20	35	45	55	55	45	45
S6	65	15	25	25		25	40	55	65	60
S7	78	20	25	25	35	45	60	60	75	85
S8	80	15	35	35		40	55	65	80	85
S9	81	15	25	30	35	50	55	65	70	65

[a]Thresholds at 1500 Hz were not obtained for all listeners.

2. Procedures

Test procedures were identical to those of experiment 1, with two exceptions. First, the S/B ratio for the test blocks was -3 dB; the S/B ratio used for the YNH listeners (-10 dB) yielded floor performance in pilot EHI subjects. Second, familiarization differed from experiment 1 in several respects. Vowel audibility was assured during familiarization by programming the TDT PF1 with an FIR filter simulating a linear hearing aid for each subject. As in experiment 1, the first familiarization block consisted of clear tokens presented in quiet with feedback. However, for the EHI listeners this training block was repeated as necessary until a criterion of 90% correct performance was met. No EHI listener required more than three presentations of the quiet familiarization block. Familiarization in noise was then carried out by presenting this familiarization block without feedback three times, at progressively more challenging S/B ratios: $+3$, 0, and -3 dB (the YNH subjects heard only one familiarization block in noise). This gradual introduction helped acclimate the EHI listeners to the babble prior to the test blocks. For the test blocks, the TDT PF1 served as a low-pass filter (8500 Hz) and attenuator to bring the signal and babble to the desired overall levels, just as in experiment 1.

B. Results and discussion

Overall intelligibility data from experiment 3 are shown in the bottom half of Table I. Surprisingly, no difference was observed between clear and conversational speech: in a two-way repeated-measures ANOVA, the main effect of speaking style was not significant [$F(1,8)=0.22$, $p=0.652$]. This result was unexpected in light of previous literature. In general, the average intelligibility advantage in clear speech for hearing-impaired listeners has ranged from 12 to 26 percentage points for sentences (Picheny *et al.*, 1985; Payton *et al.*, 1994; Schum, 1996; Uchanski *et al.*, 1996; Helfer, 1998). Figure 3 in Picheny *et al.* (1985), which shows performance for different phoneme classes, suggested an average clear speech benefit for vowels of 10 percentage points.

Methodological differences between previous studies with hearing-impaired listeners and the current experiment may explain the sharply different speaking style effects observed. One obvious difference is the identification task. All previous studies of clear speech examined sentence identification, with either verbatim or keyword scoring. However, Uchanski *et al.* (1996) also tested identification of excised key words and found a clear speech advantage of 22 percentage points. Furthermore, the results of experiment 1 demonstrate that a clear speech intelligibility advantage can be achieved for identification of vowels in excised words.

Another possibility is that the changes the current talker made in an effort to improve his intelligibility in the clear condition were not beneficial for hearing-impaired listeners. Data from the 20 talkers in Schum (1996), as well as pilot testing in this laboratory, showed that talkers vary in how much their intelligibility improves when they attempt to speak more clearly. However, when the CID Everyday Sentences were presented to the EHI subjects in noise ($+3$ dB S/B), the average clear speech benefit for this talker was 19 percentage points. This suggests that the lack of an overall clear speech advantage for vowels involved an interaction between the listeners' hearing loss and this talker's vowel production.

In contrast with the speaking style effect, the main effect of vowel and the speaking style X vowel interaction were significant [$F(9,72)=11.59$, $p<0.0001$; $F(9,72)=8.27$, $p<0.0001$]. These effects are apparent in Fig. 5 and the bottom half of Table II. The most striking result is the reversed speaking style effect observed for the front vowels.

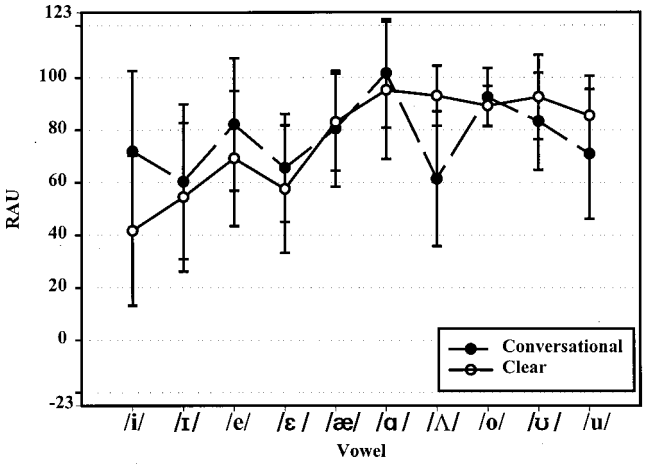


FIG. 5. Vowel identification performance in RAUs for EHI listeners. Error bars are ± 1 standard deviations from mean for nine listeners.

TABLE VI. Summary of regression analysis for EHI subjects.

	/i/	/e/	/ɛ/	/ʌ/	/ʊ/	/u/
Predictor 1	F2dis	F1dis	F2dis	F1dis	TDUR	F1dis
R^2	0.42	0.24	0.43	0.62	0.24	0.42
Predictor 2			F1dis			DM
R^2 change			0.120			0.123
Multiple R	0.65	0.49	0.74	0.79	0.49	0.73
p	<0.003	<0.03	<0.002	<0.0007	<0.03	<0.002

The average difference score (clear minus conversational) for these five vowels was −10 percentage points. For the back vowels, the average difference was positive (as expected) at 9 percentage points. A *t*-test revealed the speaking style effect for the front vowels to be significantly different from that for the back vowels ($t=-2.3$, $p=0.05$, two-tailed). These results suggest that this talker adopted a clear speech strategy that negatively affected intelligibility of front vowels for the EHI listeners.

Tables I and II also reveal that the EHI listeners had greater variance in intelligibility performance than YNH listeners, both overall and for individual vowels. This difference was particularly striking for front vowel identification. In contrast, the range in the size of the overall speaking style effect (clear minus conversational) among listeners was similar for the two groups. Across the nine YNH listeners, the range of the size of the overall speaking style effect was 14.25 (9.5 to 23.75 percentage points); the range for the nine EHI listeners was 18.25 (−8.25 to 10.0 percentage points). Thus, while every YNH listener had a clear speech advantage for vowels, the EHI subjects showed a variety of speaking style effects: only two showed a clear speech advantage, four showed essentially no difference between styles, and three had a negative clear speech effect. Within the EHI group, the size of the speaking style effect was unrelated to subject age or audiometric data ($|r|<0.3$).

VI. EXPERIMENT 4: CORRELATIONAL ANALYSIS FOR EHI LISTENERS

In experiment 4, the regression analyses from experiment 2 were replicated with data from the EHI group; the BT for each vowel was the most intelligible token for EHI listeners.

A. Method

Acoustic metrics used in these analyses were the same as in experiment 2. F1dis, F2dis, the DM, and TDUR were again the predictors; the dependent variable was the percent correct token score, converted to RAUs. Forward stepwise regression procedures were supplemented with best subset regression analyses (Neter *et al.*, 1996). The following seven vowels were selected for analysis based on criteria described in experiment 2: /i/, /e/, /ɛ/, /æ/, /ʌ/, /ʊ/, and /u/.

B. Results and discussion

Regression results for the EHI subjects are summarized in Table VI. Comparing Tables IV and VI shows that patterns of cue use for the EHI subjects differed sharply from those seen in the YNH data. The most striking difference occurred for the front vowel /æ/: while acoustic characteristics accounted for 87% of the variance in YNH intelligibility for this vowel, for the EHI subjects no acoustic factor accounted for enough variance to enter the regression equation. Raw correlations between acoustic cues and intelligibility never exceeded 0.2.

Another surprising result was the emergence of F2dis as a primary predictor of intelligibility for two of the four front vowels analyzed (/i/ and /ɛ/); F2dis was never a primary cue in the YNH group. The relationship between F2 and intelligibility for /i/ in the EHI listeners is suggested by Fig. 6. For this vowel, the raw correlation between F2 and token intelligibility was −0.68. Along the F2 dimension, conversational tokens of /i/ were closer to the BT (which was itself a conversational token) than the clear tokens, which were 30 percentage points *less* intelligible than conversational tokens. Cue use for /e/ also differed for the two groups: while YNH subjects relied primarily on TDUR, the main cue for the EHI subjects was F1dis. Finally, regression results for back vowels differed somewhat for the EHI and YNH subjects. For YNH subjects, F1dis was the main predictor of intelligibility for back vowels; this pattern was observed in the EHI subjects for /ʌ/ and /u/, but the main predictor for /ʊ/ was TDUR.

To confirm that changes in predictor importance reflect real changes and not random fluctuations inherent in the regression procedure, a jackknife analysis was performed (Efron, 1982) using the YNH data for two vowels. For each

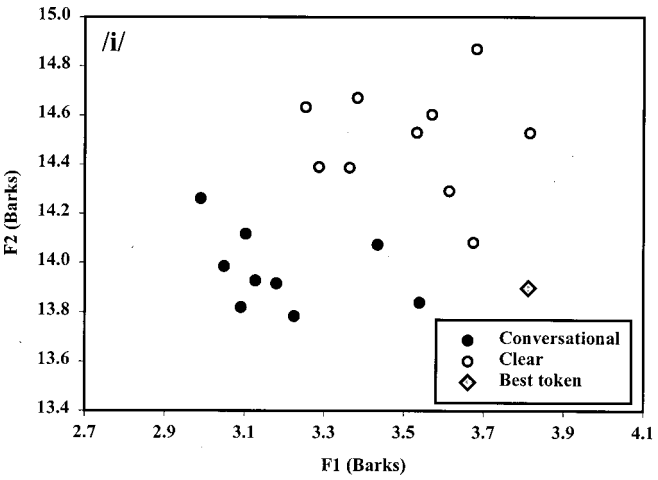


FIG. 6. Steady-state F1 and F2 frequencies for individual tokens of /i/. "Best token" refers to the most intelligible token of /i/ based on EHI intelligibility data.

vowel, the forward stepwise regression analysis was repeated for nine subsets of the subject group ($n=8$ for each subgroup). Of the 18 subgroup analyses, in only one case did the regression result differ from that obtained from the original subject group. This suggests that the forward stepwise regression procedures were robust, and that the probability is quite low that predictor order differences between conditions are random. This strongly suggests that as groups, EHI and YNH subjects made very different use of the available acoustic information when identifying vowels in noise.

VII. GENERAL DISCUSSION

Discovering the acoustic factors responsible for the generally superior intelligibility of clear speech is challenging because clear and conversational speech vary on a wide range of acoustic and phonetic dimensions (Picheny *et al.*, 1986). The approach of the current study has been to narrow the inquiry by focusing on a single phoneme class, vowels, produced by a single talker. This approach also permitted a broader exploration of the relationship between acoustic cues and vowel identification for normal-hearing and hearing-impaired listeners.

A. Summary of major results

(1) While a positive speaking style effect for natural vowels identified in noise was observed for YNH subjects, EHI subjects on average did not benefit from clear speech as produced by the talker in this study.
(2) Clear and conversational tokens of vowels differed significantly on a number of acoustic measures. Of the 10 vowels, all showed significant speaking style effects for duration along with at least one spectral metric.
(3) The speaking style manipulation generated vowel stimuli that varied in intelligibility. Token intelligibility varied sufficiently, for most vowels, to permit the use of regression analyses to establish the relative importance of three types of acoustic information to vowel identification. Our analyses suggested that steady-state formant information, dynamic formant movement, and duration all play a primary role in vowel intelligibility for normal-hearing listeners when speaking style is manipulated. This contrasts with the traditional view of duration as a secondary cue to vowel identity.
(4) Regression analyses of data from the EHI listeners indicate that hearing loss alters the way in which acoustic cues are used to identify vowels.

B. Acoustic differences between vowels in clear and conversational speech

Significant acoustic differences were found between the clear and conversational tokens of each vowel. For example, clear speech tokens had higher F1 values than conversational tokens for all vowels. This pattern was observed for one of the three talkers in Picheny *et al.* (1986) and in data averaged over eight talkers in Johnson *et al.* (1993). The effect of speaking style on F2 frequency varied, with front vowels having higher F2 in clear speech and back vowels having lower F2 in clear speech. This pattern was seen in the average data in Johnson *et al.* (1993) and for all three talkers in Picheny *et al.* (1986). Moon and Lindblom (1994) also found a similar result for the front vowels they studied. These results contrast with Ladefoged *et al.* (1976), who found no systematic F1 and F2 frequency differences among several speaking styles.

Duration was by far the most robust acoustic effect in the current study, with clear tokens averaging twice as long as conversational tokens. Vowel durations for both speaking styles were longer than those reported previously (Picheny *et al.*, 1986; Moon and Lindblom, 1994). Clear vowel tokens were more than twice as long in this study as in earlier reports, while conversational tokens were roughly 1.5 times longer. One reason for the disparity may involve differences between speech materials. For example, the materials in Moon and Lindblom (1994) were front vowels in a /wVl/ context, and only the steady-state portion of the vowel was measured. Differences among talkers may also underlie the duration differences between the current and previous studies. Not only were overall vowel durations longer than in previous studies, but also the duration difference between the two speaking styles was larger. In the current study, vowels were 2.1 times longer, on average, in clear speech than in conversational speech. For Picheny *et al.* (1986) and Moon and Lindblom (1994), clear vowels were roughly 1.3 times longer than conversational vowels. These contrasting results suggest that different talkers may adopt a variety of strategies when attempting to speak more clearly.

Measures of dynamic formant movement have not been reported in previous investigations of clear versus conversational speech or of vowel production in general. Thus, it is not known whether the degree of dynamic formant movement observed for the talker in this project is typical or atypical. Vector length, the formant movement measure employed here, was significantly greater in clear versus conversational speech only for the more crowded regions of this talker's vowel space. This suggests that the clear speech strategies a talker uses may be adapted to his or her normal speech production habits, with a goal of maximizing distinctiveness. The generality of this observation will be assessed in future studies using a newly recorded multitalker database of clear and conversational speech (Hargus Ferguson and Kewley-Port, 2002).

C. Intelligibility of vowels in clear and conversational speech

1. Between-group speaking style effect differences

While the YNH subjects enjoyed a significant clear speech vowel intelligibility advantage in experiment 1, no overall clear speech benefit was found for the EHI group in experiment 3. This discrepancy between groups was unexpected based on the previous literature. In both Payton *et al.* (1994) and Uchanski *et al.* (1996), the intelligibility advantage in clear speech for hearing-impaired listeners was similar to that achieved by normal-hearing listeners in noise.

There are two key differences between experiments 1 and 3, apart from the hearing status of the subjects, that

could have led to this discrepancy between the current and previous data. First, the two experiments used different S/B ratios (-10 and -3 dB, respectively). It is possible that the S/B ratio used for the EHI subjects was not severe enough for a clear speech effect to emerge. A second difference which yields a contrasting explanation is the age of the subjects. Many studies have shown that elderly listeners are more affected by adverse listening conditions than younger subjects are, even when effects of hearing loss have been controlled (e.g., Divenyi and Haupt, 1997; Hargus and Gordon-Salant, 1995). Pertaining to vowel perception, Coughlin *et al.* (1998) and Neel (1998) both found significantly poorer vowel identification in elderly versus young listeners. Thus, it could be that vowel perception was so degraded for the EHI subjects that no benefit could be achieved from clear speech. However, the similarity between the YNH and EHI groups for vowel identification in conversational speech (71.6% vs 73.8%, respectively) argues against either of these explanations.

2. Between-vowel speaking style effect differences

Individual vowel data for each group suggest that hearing status differences, rather than disparities in age or test condition, underlie the contrasting overall speaking style effects observed for the YNH and EHI groups. Figures 1 and 5 illustrate significant interactions between vowel and speaking style for each group and reveal very different patterns. While the magnitude of the clear speech advantage varied across vowels for both groups, for the EHI group the *direction* of the effect also varied. In general, the expected clear speech advantage was observed for the back vowels, but a *reversed* speaking style effect (clear tokens *less* intelligible than conversational tokens) occurred for the front vowels. This front/back disparity, which yields an overall noneffect of speaking style in the EHI subjects, suggests an interaction between the EHI subjects' hearing loss and the acoustic characteristics of vowels in clear and conversational speech. Figures 1 and 5 also reveal a front/back disparity in within-group intelligibility variance. While for the back vowels, variance was roughly the same for the EHI and YNH groups, for the front vowels, variance was much higher among the EHI listeners. This further suggests that front vowels were differentially affected by hearing loss.

For the front vowels, steady-state F2 frequencies were higher in clear speech than in conversational speech. The average F2 value for /i/, /ɪ/, /e/, /ɛ/, and /æ/ was 2202 Hz in clear speech and 2025 Hz in conversational speech. These values fall in a frequency region where the EHI subjects had sloping hearing loss (see Table V). Tokens with lower F2 values (conversational tokens) thus fall in a region with relatively less hearing loss than tokens with higher F2 values (clear tokens). Accordingly, conversational tokens would be more audible for the EHI subjects, and hence more intelligible than clear tokens. Negative correlations observed between F2 and token intelligibility for /i/, /ɪ/, /e/ and /ɛ/ support this assertion.

D. Relative importance of steady-state formant, dynamic formant, and duration information to vowel intelligibility

The regression analyses were run, and are interpreted in the discussions that follow, in terms of normal acoustic variability observed within individual vowel categories, as produced by a single talker. The reader should keep in mind, however, that the listeners' task was a between-category identification task. It was assumed that the most intelligible token of a given vowel category bore a closer resemblance, acoustically, to the subjects' internal representation of that category than any other category.

1. Cue use in normal-hearing listeners

In the experiment 2 regression analyses, the relative importance of steady-state target, dynamic formant, and duration information for the YNH listeners depended on the vowel. Dynamic information emerged as the most important cue for /ɛ/, duration was primary for /e/ and /æ/, and steady-state information was primary for the back vowels /ʌ/, /o/, /ʊ/, and /u/. The emergence of steady-state and dynamic information as primary cues is consistent with previous vowel perception data demonstrating the necessity of both types of cues for identifying English vowels. When only steady-state or only dynamic information has been available, vowels have been identified with approximately 75% accuracy (e.g., Hillenbrand and Gayvert, 1993). With both sources of information available, performance typically has exceeded 90% accuracy (e.g., Hillenbrand and Nearey, 1999), even in the absence of duration information (Neel, 1998).

Whereas steady-state and dynamic formant information were expected to be primary cues to vowel identity, the emergence of duration as a primary cue for even a few vowels was unexpected in light of previous data demonstrating the secondary role of duration in vowel identification. For example, in Hillenbrand *et al.* (1995), adding duration to F1 and F2 steady-state information improved the vowel classification performance of their quadratic discriminant analysis by only about 8 percentage points. The improvement achieved by adding duration was even smaller when F1 and F2 were measured at two time locations (thus providing dynamic formant information). Similarly, Strange (1989) saw silent-center syllable identification performance decline just 8 percentage points when duration information was neutralized. The only identification study where duration was manipulated independently of other parameters was carried out by Neel (1998), who found that eliminating duration reduced performance only 5 percentage points.

While the results of experiment 2 are intriguing, no generalizations can yet be made regarding the relative importance of steady state, dynamic formant, and duration information to vowel identification in normal-hearing listeners. These analyses clearly need to be replicated using materials from other talkers. For example, it is possible that listeners simply rely on the acoustic characteristics that vary the most between conversational and clear speech, and that the patterns of cue use would depend on the acoustic strategies employed by a particular talker. There is some evidence of this in the current data: the clear-conversational difference

for F1 was largest for the back vowels; accordingly, F1 was the primary cue to identity for these vowels. This argument is weakened, however, by the relatively small percentage of variance accounted for by duration overall, despite extremely large between-style duration differences. The database described above (Hargus Ferguson and Kewley-Port, 2002) will help us determine the extent to which the relationship between acoustic characteristics and vowel identity is talker dependent.

2. Effects of hearing loss on cue use

Questions of talker dependence aside, the contrasting results of experiments 2 and 4 strongly suggest that hearing loss changes the way listeners use acoustic information to identify vowels. For most of the vowels analyzed, the pattern of predictors for intelligibility differed for the YNH and EHI listener groups. For the front vowels /i/ and /ɛ/, the main predictor of intelligibility for the EHI subjects was F2; the main cue for the YNH subjects for /ɛ/ was dynamic formant movement. The EHI subjects' audiogram and the acoustic data suggest that the change in cue use for /ɛ/ was related to F2 audibility. Recall that F2 values for the front vowels fell in a frequency range where the EHI subjects had sloping hearing loss. Indeed, for the front vowels, F2 was negatively correlated with intelligibility for the EHI subjects: tokens with lower F2 values (i.e., conversational tokens) were more intelligible than tokens with higher F2 values (i.e., clear tokens). It appears that, for /ɛ/, the presence or absence of audible F2 information had more impact on intelligibility for the EHI subjects than specific patterns of formant movement.

For other vowels, there was a mixed picture. For example, for /ʌ/ and /u/, both YNH and EHI subjects enjoyed a significant clear speech intelligibility advantage, and both relied primarily on steady-state F1. Examination of the remaining vowels analyzed for both groups, (/e/, /æ/, and /ʊ/) suggests that altered cue use patterns underlie the lack of clear speech benefit observed for the EHI subjects for several vowels. While YNH subjects showed a significant clear speech benefit for /æ/ and /ʊ/, the EHI subjects showed no speaking style effect for these vowels, and regression results for the two groups differed. Similarly, for /e/, YNH and EHI subjects had very different speaking style effects (a significant positive effect for the YNH group and a significant negative effect for the EHI group) as well as very different regression results.

Taken together, the data suggest that the clear speech strategies that the current talker applied to vowels were not well suited to the vowel perception strategies of the EHI listeners. In the case of front vowels, expanding the F2 dimension of the vowel space in clear speech actually reduced intelligibility by making F2 information less audible. For other vowels, the strategies used by the current talker led to improved vowel intelligibility for the YNH subjects but had no impact on intelligibility for the EHI subjects. This is especially true of his strategy of doubling vowel duration in clear speech, which obviously did not improve intelligibility for front vowels and is thus impugned as a beneficial strategy.

E. Implications for clear speech research

This project sheds light on the acoustic features that underlie the intelligibility of clear speech by focusing on a single phoneme class. As discussed above, the changes our talker made when producing intentionally clear speech did not necessarily improve vowel intelligibility for hearing-impaired listeners. On the other hand, there was an overall clear speech benefit for sentences, so perhaps talker strategies differentially affect the intelligibility of consonants and vowels. This begs the question of whether the strategies our talker employed are unique to him, or whether they represent a typical response to instructions to speak as though talking to a hearing-impaired listener.

To date, published acoustic analyses of clear speech exist for only a handful of talkers. Comparisons of our talker with those in Picheny *et al.* (1989) and Moon and Lindblom (1994) suggest that talkers differ in the strategies they use when instructed to produce clear speech. Clear speech perceptual studies indicate that not all talkers actually achieve improved intelligibility when instructed to speak clearly (Gagné *et al.*, 1994; Schum, 1996). Similar variability in the acoustics and intelligibility of clear speech has been observed in pilot studies in this laboratory. This variability calls into question a widely held assumption, implied in previous studies and in the current project, that clear speech is a specific speaking style that automatically occurs when a talker is told to speak as though conversing with a hearing-impaired person. Talkers apparently respond in various ways to this instruction, with correspondingly varied results for intelligibility. Given that intelligibility varies among talkers even within citation-style speech (e.g., Bradlow *et al.*, 1996; Cox *et al.*, 1987), it is not surprising that variability would also be observed in clear speech, although the relationship between intelligibility performance in these two styles is unknown.

Studying clear speech can address a variety of goals. When the goal is to characterize what talkers actually do in conversations with impaired listeners, or to develop methods for training talkers to be more intelligible, a detailed understanding of talker differences in relation to improved intelligibility is required. Moreover, if the goal is to design real-time signal processing algorithms to enhance speech intelligibility, it is important to explore whether and how such strategies interact with individual talker characteristics. It is a significant outcome of the present research that some strategies employed by individual talkers to produce clear speech enhance intelligibility while others are detrimental. More research is needed to quantify the properties of clear speech that will assure benefit to persons with hearing impairment.

ACKNOWLEDGMENTS

This research was supported by NIHDCD-02229, NIHDCD-00012, and the Ira M. Ventry/Brad W. Friedrich Memorial Funds of the American Speech-Language-Hearing Foundation. The authors gratefully acknowledge the contributions of Larry E. Humes, Nicholas Hipskind, Melissa Coy-Branam, Aaron P. Brown, and Peggy Nelson, and helpful comments on a previous version from Rosalie Uchanski.

[1]Response alternatives were selected with two goals: (1) to demonstrate multiple possible spellings of each phoneme, thus emphasizing phoneme identification; and (2) to be appropriate for vowel identification in both /bVd/ and mixed CVC contexts.

Bradlow, A., Torretta, G., and Pisoni, D. (**1996**). ''Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics,'' Speech Commun. **20**, 255–272.

Coughlin, M., Kewley-Port, D., and Humes, L. (**1998**). ''The relation between identification and discrimination of vowels in young and elderly listeners,'' J. Acoust. Soc. Am. **104**, 3597–3607.

Cox, R., Alexander, G., and Gilmore, C. (**1987**). ''Intelligibility of average talkers in typical listening environments,'' J. Acoust. Soc. Am. **81**, 1598–1608.

Davis, H., and Silverman, S. (**1978**). *Hearing and Deafness*, 4th ed. (Holt, Rinehart and Winston, New York).

Divenyi, P., and Haupt, K. (**1997**). ''Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. I. Age and lateral asymmetry effects,'' Ear Hear. **18**, 42–61.

Efron, B. (**1982**). *The Jackknife, the Bootstrap, and Other Resampling Plans* (Society for Industrial and Applied Mathematics, Philadelphia, PA).

Gagné, J.-P., Masterson, V., Munhall, K., Bilida, N., and Querengesser, C. (**1994**). ''Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech,'' J. Acad. Rehabil. Audiol. **27**, 135–158.

Gagné, J.-P., Querengesser, C., Folkeard, P., Munhall, K., and Masterson, V. (**1995**). ''Auditory, visual, and audiovisual speech intelligibility for sentence-length stimuli: An investigation of conversational and clear speech,'' The Volta Review **97**, 33–51.

Hargus, S., and Gordon-Salant, S. (**1995**). ''Accuracy of Speech Intelligibility Index predictions for noise-masked young listeners with normal hearing and for elderly listeners with hearing impairment,'' J. Speech Hear. Res. **38**, 234–243.

Hargus Ferguson, S., and Kewley-Port, D. (**2002**). ''Vowels in clear and conversational speech: Talker difference in acoustic characteristics and intelligibility for normal-hearing listeners,'' J. Acoust. Soc. Am. **111**, 2482.

Helfer, K. (**1998**). ''Auditory and auditory-visual recognition of clear and conversational speech by older adults,'' J. Am. Acad. Audiol. **9**, 234–242.

Hillenbrand, J., and Gayvert, R. (**1993**). ''Identification of steady-state vowels synthesized from the Peterson and Barney measurements,'' J. Acoust. Soc. Am. **94**, 668–674.

Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Hillenbrand, J., and Nearey, T. (**1999**). ''Identification of resynthesized /hVd/ utterances: Effects of formant contour,'' J. Acoust. Soc. Am. **105**, 3509–3523.

Johnson, K., Flemming, E., and Wright, R. (**1993**). ''The hyperspace effect: Phonetic targets are hyperarticulated,'' Language **69**, 505–528.

Kalikow, D., Stevens, K., and Elliott, L. (**1977**). ''Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability,'' J. Acoust. Soc. Am. **61**, 1337–1351.

Kewley-Port, D., Akahane-Yamada, R., and Aikawa, K. (**1996**). ''Intelligibility and acoustic correlates of Japanese accented English vowels,'' Proc. ICSLP 96, pp. 450–453.

Ladefoged, P., Kameny, I., and Brackenridge, W. (**1976**). ''Acoustic effects of style of speech,'' J. Acoust. Soc. Am. **59**, 228–231.

Liénard, J.-S., and DiBenedetto, M.-G. (**1999**). ''Effect of vocal effort on spectral properties of vowels,'' J. Acoust. Soc. Am. **106**, 411–422.

Loizou, P. (**2000**). COLEA: A Matlab Software Tool for Speech Analysis. Available: http://www.utdallas.edu/~loizou/speech/colea.htm (Accessed 3/17/99).

Moon, S.-J., and Lindblom, B. (**1994**). ''Interaction between duration, context, and speaking style in English stressed vowels,'' J. Acoust. Soc. Am. **96**, 40–55.

Neel, A. (**1998**). Factors influencing vowel identification in elderly hearing-impaired listeners, Ph.D. dissertation, Indiana University.

Nejime, Y., and Moore, B. (**1998**). ''Evaluation of the effect of speech-rate slowing on speech intelligibility in noise using a simulation of cochlear hearing loss,'' J. Acoust. Soc. Am. **103**, 572–576.

Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. (**1996**). *Applied Linear Statistical Models*, 4th ed. (Times Mirror Higher Education Group, Chicago).

Payton, K., Uchanski, R., and Braida, L. (**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.

Picheny, M., Durlach, N., and Braida, L. (**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., Durlach, N., and Braida, L. (**1986**). ''Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech,'' J. Speech Hear. Res. **29**, 434–446.

Picheny, M., Durlach, N., and Braida, L. (**1989**). ''Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to difference in intelligibility between clear and conversational speech,'' J. Speech Hear. Res. **32**, 600–603.

Schum, D. (**1996**). ''Intelligibility of clear and conversational speech of young and elderly talkers,'' J. Am. Acad. Audiol. **7**, 212–218.

Statistica for Windows [Computer software] (**1999**). StatSoft, Inc., Tulsa, OK.

Strange, W. (**1989**). ''Dynamic specification of coarticulated vowels spoken in sentence context,'' J. Acoust. Soc. Am. **85**, 2135–2153.

Studebaker, G. (**1985**). ''A ''rationalized'' arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.

Summers, W., and Leek, M. (**1992**). ''The role of spectral and temporal cues in vowel identification by listeners with impaired hearing,'' J. Speech Hear. Res. **35**, 1189–1199.

Traunmüller, H. (**1990**). ''Analytical expressions for the tonotopic sensory scale,'' J. Acoust. Soc. Am. **88**, 97–100.

Uchanski, R., Choi, S., Braida, L., and Durlach, N. (**1996**). ''Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate,'' J. Speech Hear. Res. **39**, 494–509.

Wechsler, D. (**1983**). *Wechsler Adult Intelligence Scale—Revised* (The Psychological Corporation, New York).

The clear speech effect for non-native listeners[a)]

Ann R. Bradlow[b)] and Tessa Bent
Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208

(Received 24 July 2001; accepted for publication 24 April 2002)

Previous work has established that naturally produced clear speech is more intelligible than conversational speech for adult hearing-impaired listeners and normal-hearing listeners under degraded listening conditions. The major goal of the present study was to investigate the extent to which naturally produced clear speech is an effective intelligibility enhancement strategy for non-native listeners. Thirty-two non-native and 32 native listeners were presented with naturally produced English sentences. Factors that varied were speaking style (conversational versus clear), signal-to-noise ratio (-4 versus -8 dB) and talker (one male versus one female). Results showed that while native listeners derived a substantial benefit from naturally produced clear speech (an improvement of about 16 rau units on a keyword-correct count), non-native listeners exhibited only a small clear speech effect (an improvement of only 5 rau units). This relatively small clear speech effect for non-native listeners is interpreted as a consequence of the fact that clear speech is essentially native-listener oriented, and therefore is only beneficial to listeners with extensive experience with the sound structure of the target language.
[DOI: 10.1121/1.1487837]

PACS numbers: 43.71.Gv, 43.71.Es, 43.71.Hw [KRK]

I. INTRODUCTION

Previous work has shown that naturally produced clear speech is more intelligible than conversational speech for a variety of listener populations under various listening conditions. In a seminal paper, Picheny, Durlach, and Braida (1985) reported a 17–20-percentage point clear speech intelligibility advantage for hearing-impaired adults. This clear speech effect was robust across several talkers, listeners, presentation levels, and frequency-gain characteristics. This result has since been extended to both normal and hearing-impaired listeners in noise and reverberation (Payton *et al.*, 1994; Uchanski *et al.*, 1996), and to elderly hearing-impaired listeners (Schum, 1996). While there is some variance across talkers in the magnitude of the clear speech intelligibility advantage they can provoke, the available data indicate that the clear speech effect (i.e., the intelligibility difference between conversational and clear speech) is generally robust across a wide range of talkers, listeners, and signal degradation conditions.

The non-native listener population is in many respects the opposite of the listener population that has typically been the focus of clear speech research. Most previous studies of clear speech perception have focused on communicative situations in which speech communication is challenged by the introduction of some factor that impedes the listener's access to the signal either by signal degradation (such as the addition of background noise or reverberation) or by the listener's own hearing loss. In contrast to the problems of signal access experienced by normal listeners in noise and reverberation or by hearing-impaired listeners, the underlying source of the speech perception problems experienced by non-native listeners is their limited experience with the sound system (i.e., the system of phonological contrasts and their phonetic realization) and higher levels of linguistic structure (i.e., vocabulary, syntax, semantics, and pragmatics) of the target language. In other words, non-native listener speech perception deficits arise primarily from problems of access to the language-specific, linguistic code, rather than from problems of access to the speech signal.

The major goal of the present study was to investigate the extent to which naturally produced clear speech is an effective intelligibility enhancement strategy for non-native listeners. If clear speech enhances speech perception in the face of either signal or code access problems (that is, clear speech is effective for both hearing-impaired and non-native listeners), then we may conclude that the intelligibility advantage of clear speech is not limited to listeners who already have extensive experience with the target language. However, if naturally produced clear speech is not as effective a speech perception enhancement strategy for non-native listeners as for native listeners with signal access problems, then we may conclude that clear speech production is essentially native-listener oriented, and its perceptual benefits are only accessible to listeners who already have extensive experience with the target language. Either of these two results would be of considerable interest from both theoretical and practical perspectives since it would provide insight into the mechanism that underlies the intelligibility advantage of clear speech and suggest directions to follow for the design of speech perception improvement strategies that can be optimized for specific listener populations.

Previous research on clear speech production has identified a wide range of acoustic phonetic features that characterize the conversational-to-clear speech transformation. These include modifications that serve to enhance the overall acoustic salience of the signal such that it is more resistant to the adverse effects of background noise or a listener-related perceptual deficit. These modifications include a decreased

[a)]An earlier version of this study was presented at the 141st meeting of the Acoustical Society of America, Chicago, IL, June 2001.
[b)]Electronic mail: abradlow@northwestern.edu

speaking rate (including the insertion of longer and more frequent interword pauses), a wider dynamic pitch range, greater sound-pressure levels, more salient stop consonant releases, and greater obstruent rms intensity (Picheney *et al.*, 1986).

Additionally, clear speech production involves modifications that serve to enhance the acoustic distance between contrasting speech categories. For example, clear speech production involves less vowel reduction, thereby enhancing the acoustic distance between individual vowels (Picheney *et al.*, 1986; Moon and Lindblom, 1994; Johnson, Flemming, and Wright, 1993; Bradlow, in press; Ferguson and Kewley-Port, in press). Furthermore, Uchanski (1988, 1992) showed that clear speech production involves nonuniform segment duration increases such that durational contrasts are enhanced; for example, while both long/tense and short/lax vowels increased in duration from conversational to clear speech, the lengthening was far greater for the long/tense vowels than for the short/lax vowels, thereby enhancing the length distinction between these contrasting categories. Similarly, Cutler and Butterfield (1990) showed that English syllables before word boundaries were lengthened to a greater extent in clear speech than in conversational speech, thereby enhancing the duration difference between preboundary and nonpreboundary syllables. Thus, the available data indicate that clear speech production is guided by both signal enhancement strategies (i.e., those that make the signal generally more acoustically salient) and code enhancement strategies (i.e., those that exaggerate the acoustic distance between contrasting categories).

In addition to incorporating these signal and code enhancements, clear speech production also involves the maintenance of certain language-specific pronunciation norms. For example, Bradlow (in press) showed that clear speech production in both English and Spanish involved maintenance (rather than reduction) of CV coarticulation throughout the duration of long, clear speech syllables where passive constraints on articulator movements could have been overcome (see also Matthies *et al.*, 2001). Furthermore, Ohala (1995) showed that temporal differences that distinguish English stop consonant categories, such as voice onset time and overall word duration, were not exaggerated in clear speech; rather, talkers exhibited a strict adherence to pronunciation norms of the language vis à vis these parameters during clear speech production. In combination with the findings about the enhancement of the acoustic distance between contrasting categories, these findings suggest that clear speech production is guided, at least in part, by principles that directly reflect the phonology and phonetics of the target language.

Previous research on non-native speech perception has identified two major sources of non-native listener difficulty. The first of these sources is specific to the particular native-target language (L1–L2) pairing in question; the second is apparently independent of the language backgrounds of the listener and talker. The L1–L2 specific source of non-native listener difficulty has been well documented (for example, see the contributions to Strange, 1995), and results from the fact that listeners perceive an incoming L2 speech sound as belonging to the L1 category whose members are most similar to it (Best, 1994, 1995; Flege, 1992, 1995; Kuhl and Iverson, 1995). That is, non-native speech sounds are perceived with respect to the sound system, or code, of the native language. In addition to this code mismatch between L1 and L2 which interferes with L2 speech perception, non-native listeners also appear to be disproportionately challenged by degraded signals relative to native listeners. For example, several studies have shown that non-native listener performance declines more sharply than native-listener performance on speech perception tasks with increasing levels of signal distortion either through the addition of background noise or reverberation (Nábèlek and Donahue, 1984; Mayo, Florentine, and Buus, 1997; Meador, Flege, and McKay, 2000). Thus, the difficulties imposed by the non-native listeners' lack of experience with the phonological structure and the patterns of phonetic implementation of the target language (i.e., their code access problems) are aggravated by the presence of a signal distorting influence.

Based on this information about clear speech production and about patterns of non-native listener speech perception, we predicted that non-native listeners who have had very limited experience with the spoken form of the target language would show only a small intelligibility benefit for clear speech relative to conversational speech. That is, we expected to find that clear speech would not be as effective a speech intelligibility enhancement tool for non-native listeners as it is for hearing-impaired listeners or for normal listeners with degraded speech. The rationale behind this prediction was that, while the signal enhancements of clear speech production should provide a perceptual benefit to all listeners, the code enhancements (including those that enhance the acoustic distance between contrasting categories and those that involve adherence to pronunciation norms) are likely to be beneficial only for those listeners who are already experienced with the sound structure of the target language. In particular, the slower speaking rate, the wider dynamic pitch range, greater sound pressure levels, more salient stop consonant releases, and greater obstruent rms intensity (i.e., the signal enhancements of clear speech) should promote intelligibility even for listeners who are relatively unskilled in decoding connected speech in the target language (Hazan and Simpson, 2000). However, only a listener who is already sensitive to the important dimensions of contrast that define the sound system of the target language or who is familiar with the detailed pronunciation patterns of native speakers is likely to benefit from the code enhancements of clear speech. For example, only a listener who is already sensitive to the consistent length difference between lax/short and tense/long vowels in English will benefit from the enhanced duration difference between these vowel categories in clear speech. Therefore, we predicted that overall non-native listeners would derive a significant, but relatively small, clear speech benefit.

In order to test this prediction we designed a test of sentence-in-noise perception in which we manipulated speaking style (conversational versus clear) and signal-to-noise ratio (-4 vs -8 dB). Our specific research questions were: (1) How does the magnitude of the clear speech benefit compare across native and non-native listeners? (2) How do

TABLE I. General information about the non-native subjects.

	Number of subjects reporting	Mean	Range
Age	32	25.5 yrs.	22–32 yrs.
Length of English study	31	9.8 yrs.	6–17 yrs.
Age of acquisition of English	31	12.0 yrs.	5–18 yrs.
Length of time in an English speaking community	32	0.26 yrs.	0.02–2 yrs.
TOEFL score[a]	25	626	597–667
SPEAK score[b]	27	37.5	27–52

[a]TOEFL = Test of English as a Foreign Language (mostly written).
[b]SPEAK = Speaking Proficiency English Assessment Kit (a test of spoken English).

increasing levels of noise affect sentence perception by native and non-native listeners? In addition to these primary research questions, we were also interested in investigating some of the factors that may be related to sentence-in-noise perception abilities across individual non-native listeners. Specifically, we wanted to see whether non-native listener sentence-in-noise perception ability correlated with speech production ability and with various demographic variables, including age of acquisition, length of acquisition, and written language ability (as measured by the standardized Test of English as a Foreign Language, or TOEFL).

II. METHOD

The overall design of the study involved testing a group of adult non-native listeners who had very limited prior exposure to spoken English on a sentence-in-noise recognition task. Performance on this task was measured as a function of speaking style and level of background noise. A control group of normal-hearing native American English listeners was also tested on this task. As part of this experiment, we also collected data on the non-native subjects' familiarity with the words in the test sentences, and subsequently recorded them producing the set of test sentences. These recordings were then submitted to intelligibility testing by native American English listeners, thereby allowing us to examine the correlation between sentence perception and production for the group of non-natives. This overall design yielded two primary sets of data: (1) performance on the sentence-in-noise perception task by groups of non-native and native listeners, and (2) sentence production accuracy by the same non-native subjects who performed the perception test as judged by an independent group of native listeners (i.e., not the same native listeners who participated in the perception test).

A. Subjects

A total of 64 normal-hearing adults participated in this study. The test group included 32 non-native speakers of English, recruited from the Northwestern University International Summer Institute and English as a Second Language Program. The International Summer Institute is designed to provide incoming international graduate students from across the university with intensive English language training as well as a general introduction to academic life in the USA during the month before they begin their graduate studies at Northwestern University. All of the participants in this program had already been admitted to a doctoral program and had therefore demonstrated a high level of proficiency with written English communication (as measured by a minimum score of 560 on the paper-and-pencil TOEFL examination or 220 on the computer-based version of the test). However, these students had been identified by their admitting department as likely to have some difficulty with spoken English communication (based on their department's subjective experience of the spoken English skills of previous students from each individual student's home country). Similarly, the subjects recruited from the ESL program all came to the program due to their own (or their department's) recognition of their need to improve their oral and aural English skills. Thus, the population in this study does not represent a random sampling of non-native speakers in the Northwestern University community, nor were the subjects selected based on their native language background; rather, these individuals were selected based on their limited experience in an English-speaking environment. The breakdown according to native language background was as follows: Bengali ($n=1$); Chinese ($n=20$); Hindi ($n=1$); Japanese ($n=1$); Korean ($n=5$); Romanian ($n=1$); Slovakian ($n=1$); Spanish ($n=1$); Thai ($n=1$). Additional information about the group of non-native subjects is provided in Table I. (As indicated in the table, some subjects were not able or did not wish to report some of the information we requested.) The non-native subjects were all paid for their participation in this study.

The native speakers in the control group were all currently enrolled undergraduates at Northwestern University. They ranged in age from 18–30 years, and were all monolingual speakers of American English. They were recruited from the Linguistics Department subject pool and received course credit for their participation in this study. All subjects (from both the non-native test group and the native control group) reported normal speech and hearing at the time of testing.

B. Stimuli

The stimuli for this study consisted of slightly modified versions of the sentence lists included in the Revised Bamford–Kowal–Bench Standard Sentence Test, which was originally developed for use with British children (Bench and Bamford, 1979). The revised set of sentence lists was developed by the Cochlear Corporation for use with American children. Each list consists of 16 simple English sen-

tences with either three or four keywords for a total of 50 keywords per list. Of the original 21 lists, four lists (lists 7, 8, 9, and 10) were selected based on their equivalent intelligibility scores for normal hearing children as reported in Bamford and Wilson (1979). The rationale behind the selection of these sentences for this study was that, due to their limited vocabulary, they would be appropriate for use with a variety of listener populations, including non-native listeners and children (both of whom are likely to have limited receptive and productive vocabularies). The sentence lists used in this study are provided in the Appendix.

Two adult native speakers of Standard American English with no known speech or hearing impairment (one male, age 33 years, and one female, age 40 years) were recorded producing these sentences in a sound-treated booth in the Phonetics Laboratory in the Department of Linguistics at Northwestern University. They read the sentences from a printed list, speaking into a microphone that fed directly into the sound card (SoundBlaster Live) of a desktop computer. Recording was done on a single channel at a sampling rate of 16 kHz using the PRAAT speech analysis software package (developed at The Institute of Phonetic Sciences at the University of Amsterdam, copyright by Paul Boersma and Paul Weenink). The input level was continuously monitored and adjusted so as to ensure maximum gain without exceeding the dynamic range of the recording system.

The talkers produced the sentences under two conditions. First they read the sentences in a conversational speaking style, then they repeated the sentences in a clear speaking style. For the conversational speaking style, they were instructed to read at their normal pace without any particular attention to clarity, as if addressing someone highly familiar with their voice and speech patterns. For the clear speaking style, they were instructed to read the sentences as if speaking to a listener with a hearing loss or from a different language background.

Following the recording sessions, the digital speech files were segmented into sentence-length files. The root-mean-square amplitude of each of the digital speech files was then rescaled to 65 dB SPL. Each file was then digitally mixed with white noise (generated in PRAAT) to yield two speech-plus-noise files, with −4 and −8 dB as the two signal-to-noise ratios. These signal-to-noise ratios were selected on the basis of pilot testing which indicated that normal-hearing, native listeners performed in the mid-to-high range on a sentence intelligibility test with these sentences at these signal-to-noise ratios. Each of the final stimulus files consisted of a 400-millisecond silent leader, followed by 500 milliseconds of noise, followed by the speech-plus-noise file, and ending with a 500-millisecond noise-only tail. The noise in the 500-millisecond, noise-only header and tail was always at the same level as the noise in the speech-plus-noise portion of the stimulus file, that is at 69 and 73 dB for the −4- and −8-dB signal-to-noise ratio conditions, respectively.

C. Procedure

1. Non-native listeners

Each subject participated in two data collection sessions: a perception test session followed by a production recording session. For the first session, subjects were tested in groups of one to three. The data collection session began with a language background questionnaire which probed the subjects' language learning experiences (both native and foreign languages) as well as their self-reported performance on standardized tests of English language proficiency. Next, subjects were seated in front of a computer monitor in a sound-attenuated booth and the sentence-in-noise perception test was administered. Stimulus presentation was controlled by special-purpose experiment running software (SUPERLAB PRO 2.01). The audio files were played out through the computer sound card (SoundBlaster Live) over headphones (Sennheiser HD580) at a comfortable listening level, which was set by the experimenter before the start of the experiment. The subject's task was to listen to the sentence stimulus and to write down whatever she or he heard on specially prepared answer sheets. After each trial, the subject pressed a button on a response box (supplied as part of the SUPERLAB PRO 2.01 package) to elicit the next trial. Each trial was presented only once, but subjects could take as long as they needed to record their responses.

Each subject responded to stimuli from only one talker (either the male or the female talker). Signal-to-noise ratio (−4 versus −8 dB), and speaking style (conversational versus clear) were factors that varied within subjects. Each subject responded to a total of 64 sentences (4 sets of 16 sentences), which were distributed evenly across the two signal-to-noise and speaking style conditions. For all subjects, the first two sets of sentences were presented in the −4-dB signal-to-noise ratio condition, and the second two sets were presented in the -8-dB signal-to-noise ratio condition. This order of presentation of the two signal-to-noise ratio conditions was done to ensure that any practice effect would be counteracted by the decline in signal-to-noise ratio (which was presumed to be more powerful than any practice effect). Within each of the two signal-to-noise ratio blocks, one set of sentences was presented in conversational speaking style and the other was presented in the clear speaking style. The order of presentation of the four sentence lists was counterbalanced across subjects. Over the course of the entire session, each subject heard each of the 64 sentences only once.

After all subjects had finished the sentence-in-noise perception test, a word familiarity rating test was administered. For this test, each of the 154 unique keywords in the complete set of sentences used in the sentence-in-noise perception test was presented to the subjects for a familiarity rating on a scale of 1 to 7 where 1="I don't know this word," 4="I recognize this as an English word but I don't know its meaning," and 7="I know this word." An additional set of 75 filler items was presented as part of this test. These additional words were selected from lists of words that were given low, medium, and high familiarity ratings by native listeners in Lewellen *et al.* (1993) and that were used in previous tests with both native and non-native listeners (Bradlow and Pisoni, 1999). The inclusion of these words ensured that the full range of the familiarity rating scale would be represented by the items in this test. On each trial, the target word was presented in standard American English orthography on the computer screen (using SUPERLAB PRO 2.01 soft-

ware), and the subject entered his or her familiarity rating by pressing the appropriate button on the keyboard. The current item remained on the screen until a response was recorded, which then triggered the start of the next trial. The order of presentation of the items was randomized. Both ratings and response times were recorded. Subjects were instructed to work as fast as they could without sacrificing accuracy. The entire first data collection session (including the language background questionnaire, the sentence-in-noise perception test, and the word familiarity rating test) lasted approximately 1 hour.

Subjects returned to the Phonetics Laboratory approximately 1–2 weeks after the first data collection session for a recording session. In this session, subjects were recorded individually reading the four sets of sentences that were used in the sentence-in-noise perception test. They were instructed to read at their natural pace without any particular attention to clarity and were given no indication as to the nature of the intended listener. These recordings were made on an Ariel Proport with a Shure SM81 microphone. All subjects read the four sets of sentences in the same order. After the recording, the sound files were converted to the WAV format and transferred to a PC-based computer. The digital speech files were then segmented into sentence length files, and each file's root-mean-square amplitude was rescaled to 65 dB SPL. Each file was then digitally mixed with white noise (generated in PRAAT) to yield a speech-plus-noise file with a +5-dB signal-to-noise ratio. The final files had the same composition as the sentence-in-noise test files: a 400-millisecond silent leader, followed by 500-milliseconds of noise, followed by the speech-plus-noise file, and ending with a 500-millisecond noise-only tail.

2. Native listeners

Two independent groups of native listeners participated in this study. The first group of 32 native listeners served as a control group for the sentence-in-noise perception test that was administered to the non-native listeners. Data collection on this test with the native listener controls followed exactly the same procedure as for the non-native listeners.

The second group of native listeners ($n=40$) served as judges for the non-native sentence productions that were recorded in the second session with the non-native subjects. In order to keep the number of required native listener judges manageable, each non-native talker's overall intelligibility was assessed on the basis of native listener transcriptions of only one of the four sentence lists. Furthermore, in order to avoid possible adaptation on the part of the native listener to the non-native talker's voice and speech patterns, we compiled 8 sentence presentation conditions each of which included 2 sentence productions from each of the 32 non-native talkers for a total of 64 sentences per condition. According to this scheme, each talker contributed 2 sentences to each of the 8 conditions resulting in each talker's overall intelligibility score being based on transcriptions of one set of 16 sentences (2 sentences in each of the 8 condition) from a total of 40 independent listeners (5 listeners in each of the 8 conditions).

D. Data analysis

Accuracy scores on the sentence-in-noise perception test from both the non-native and native listeners were obtained by counting the number of keywords correctly transcribed. We adopted a strict scoring criterion by which a word was counted as correctly transcribed if it included all and only the appropriate affixes (e.g., if a plural marking "s" or past tense marking "ed" was omitted or added the word was counted as incorrect.) Listeners were not penalized for obvious spelling errors. This scoring method resulted in a keyword-correct score out of a possible 50 for each of the four sentence lists for each listener: −4-dB signal-to-noise ratio conversational style, −4-dB signal-to-noise ratio clear style, −8-dB signal-to-noise ratio conversational style, −8-dB signal-to-noise ratio clear style. The scores were converted to percent-correct scores, and then converted to rationalized arcsine transform units (rau) (Studebaker, 1985). This transformation places the scores on a linear and additive scale, thus facilitating meaningful statistical comparisons across the entire range of the scale. An identical scoring system was applied to the native listener transcriptions of the non-native productions.

It is important to note that both the use of meaningful sentences and the strict scoring criterion (in which all and only the appropriate affixes had to be present for a word to be counted as correct) may have penalized the non-native listeners based on their more limited knowledge of English morphology and syntax relative to the native listener controls. Therefore, while the data from this experiment do not allow us to view the effect of language background (native versus non-native) on speech perception in the absence of higher-level linguistic influences on sentence recognition accuracy, they provide a benchmark for assessing sentence recognition under conditions that, at least to some extent, resemble real-world communicative situations. It remains for further research to compare performance under the sentence and data analysis conditions of the present experiment with performance under conditions that isolate speech perception ability from other higher-level linguistic influences.

III. RESULTS

Data from the word familiarity rating task with the non-native listeners showed that the vast majority of the words was highly familiar to the vast majority of the listeners. Familiarity rating data from 2 of the 32 listeners had to be discarded due to an equipment failure in one case and a misunderstanding of the task on the subject's part in the other case. Of the remaining 30 subjects, 6 gave high ratings (greater than 5 on the 7-point scale) to all of the 154 unique keywords in the sentence-in-noise perception test. Of the remaining 24 listeners, only 2 gave low ratings (less than 5) to more than 4% of the words: 1 gave low ratings to 7 words (4.5%) and 1 gave low rating to 11 words (7%). The rest of the listeners gave high ratings (greater than 5) to at least 97% of the words. We found no relationship between the number of low-familiarity ratings given by an individual subject and his or her overall performance on the sentence-in-noise perception test. Therefore, we performed all analyses of the

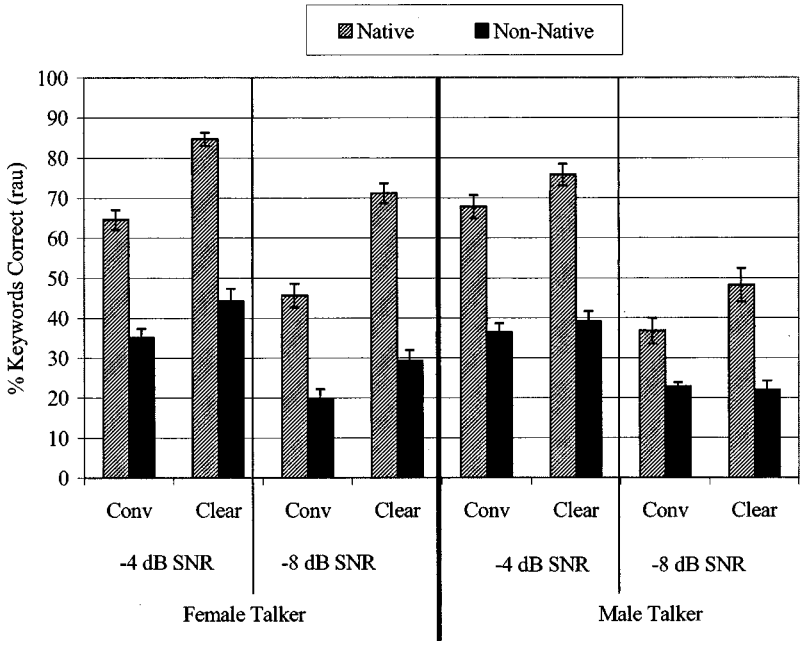


FIG. 1. Overall sentence-in-noise perception test scores (mean keywords correct on the rau scale) for native and non-native listeners. Error bars are the standard error of the mean.

sentence-in-noise perception test with the assumption that the non-native listeners were all sufficiently familiar with the keywords to ensure that this test provided a valid measure of their ability to perceive sentences in noise under the various speaking styles and signal-to-noise ratio conditions incorporated into this test.

Figure 1 shows the overall results of the sentence-in-noise perception test for both non-native and native listeners. Data for subjects that responded to the stimuli from the female and male talkers are shown on the left and right, respectively. As expected, the non-native listeners as a group performed at a much lower level on this test than the native listeners. Furthermore, the two groups of subjects showed different patterns of responses across the various conditions. The data in absolute rau units were submitted to a four-factor repeated measures ANOVA with talker (male versus female) and listener (native versus non-native) as between-subjects factors, and signal-to-noise ratio (-4 vs -8 dB), and speaking style (conversational versus clear) as within-subjects factors. All four main effects were significant [talker: $F(1,60) = 9.23$, $p < 0.005$; listener: $F(1,60) = 272.64$, $p < 0.0001$; signal-to-noise ratio: $F(1,60) = 248.75$, $p < 0.0001$; speaking style: $F(1,60) = 93.32$, $p < 0.0001$]. Listener interacted significantly with speaking style [$F(1,60) = 25.08$, $p < 0.0001$] due to the fact that the native listeners exhibited a much greater clear speech effect than the non-native listeners [mean difference $= 11.11$ rau; $t(62) = 4.28$, $p < 0.0001$]. The speaking style by talker interaction was also significant [$F(1,60) = 23.66$, $p < 0.0001$] due to the fact that the female talker elicited a much greater clear speech effect than the male talker [mean difference $= 10.79$ rau; $t(62) = 4.125$, $p < 0.0001$]. Signal-to-noise ratio entered into significant three-way interaction with talker and listener [$F(1,60) = 7.49$, $p < 0.01$] due to the fact that, for the native listeners, the decline in performance for the -8-dB signal-to-noise ratio condition relative to the -4-dB signal-to-noise ratio condition was greater for the stimuli produced by the male talker than for the stimuli produced by the female talker [mean difference $= 13$ rau; $t(30) = 3.47$, $p < 0.001$], whereas for the non-native listeners there was no significant difference in the signal-to-noise ratio effect across the two talkers [mean difference < 1 rau; $t(30) = 0.06$, $p = 0.96$]. None of the other interactions was significant.

Table II shows the proportional changes that correspond to the absolute scores shown in Fig. 1. For the improvement from conversational to clear speech perception, these proportions were calculated as (clear speech score−conversational speech score)/conversational speech score, and therefore represent the clear speech benefit as a proportion of the conversational speech score. For the decline in performance from the -4-dB signal-to-noise ratio condition to the -8-dB signal-to-noise ratio condition, these proportions were calculated as $(-4\text{ dB})–(-8\text{ dB})/(-4\text{ dB})$, and therefore represent the decline in speech perception performance as a proportion of performance in the -4-dB signal-to-noise ratio condition.

Figure 2 shows the magnitude of the clear speech effect (top panels) and of the signal-to-noise ratio effect (bottom panels) for individual subjects in the male (left panels) and female (right panels) talker conditions. The clear speech effect sizes shown in this figure were calculated by subtracting

TABLE II. Proportional changes in sentence-in-noise performance.

		Native	Non-native
Proportional increase (Clear-Conv)/Conv:			
Female	−4 dB	31%	26%
	−8 dB	56%	49%
Male	−4 dB	12%	7%
	−8 dB	31%	−4%
Proportional decrease (−4 dB)−(−8 dB)/(−4 dB):			
Female	Clear	16%	34%
	Conv.	29%	44%
Male	Clear	36%	44%
	Conv.	46%	37%

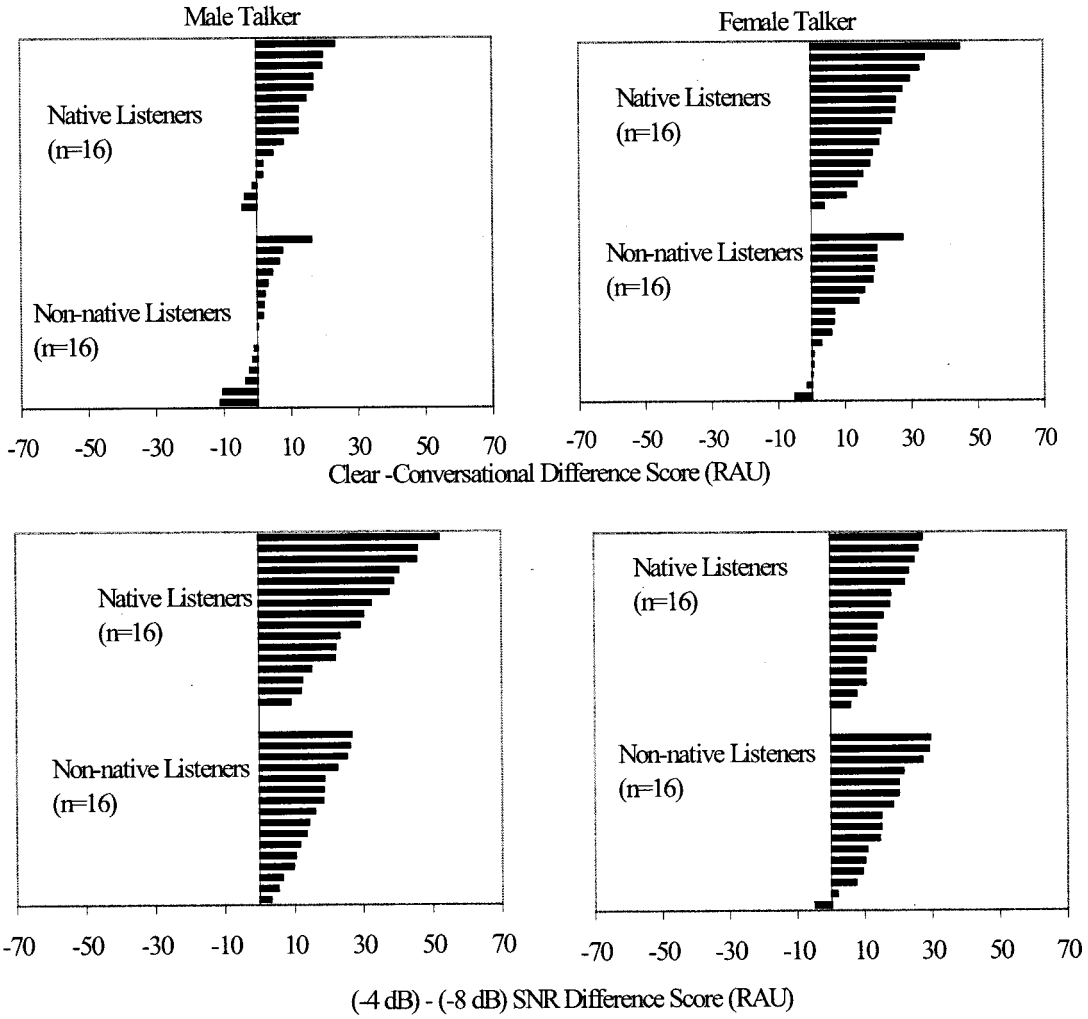


FIG. 2. Magnitude of the clear speech effect (top panels) and of the signal-to-noise ratio effect (bottom panels) for individual subjects in the male (left panels) and female (right panels) talker conditions.

each listener's average sentence-in-noise perception score (across the two signal-to-noise ratio conditions) for the conversational speech stimuli from the corresponding score for the clear speech stimuli. Similarly, the signal-to-noise ratio effect sizes were calculated by subtracting each listener's average sentence-in-noise perception score (across the two speaking style conditions) for the −8-dB signal-to-noise ratio from the corresponding score for the −4-dB signal-to-noise ratio. The plots show that the group mean effects, discussed above and shown in Fig. 1, are reflected quite well when we look at the individual subject data.

In order to gain some insight into the factors that may have contributed to some of the individual variability observed in the size of the clear speech effect across individual non-native listeners, we looked at the relationship between the sentence-in-noise perception scores and the various demographic variables that were collected in the language background questionnaire. This analysis was performed using various sentence-in-noise perception scores: only the conversational speech stimuli, only the clear speech stimuli, and the clear-conversational speech difference score. The most notable result of this correlational analysis (given in Table III) was a negative correlation ($r=-0.358$, $p<0.05$) between conversational speech perception score and the magnitude of the clear speech effect, indicating that those non-native listeners with poorer conversational speech perception benefit more from clear speech than their fellow non-native listeners who begin with better conversational speech perception. There was also a significant positive correlation between clear speech perception scores and the magnitude of the clear speech effect ($r=+0.738$, $p<0.05$), probably because relatively good performance with clear speech typically arises from a relatively large clear speech effect. Also of note in this correlational analysis was the positive correlation between sentence production (as assessed by the native listeners' transcriptions of the non-native talkers' productions) and clear speech perception ($r=+0.449$, $p<0.01$), indicating that those non-natives with relatively good perception abilities also had relatively good production abilities. This perception–production correlation was weaker and just failed to reach significance when perception ability was indexed by conversational speech perception ability ($r=+0.34$, $p<0.0564$), possibly due to the fact that the conversational speech perception scores showed less variance than the clear speech perception scores (standard deviation for conversational and clear speech perception were 6.67 and 9.23 rau, respectively).

TABLE III. Correlations between sentence-in-noise perception scores and various other demographic variables for the non-native listeners.

	Conv.	Clear	Clear-Conv.	TSE	TOEFL	AOA	LOA	LOI
Conversational ($n=32$, 5–42 rau)								
Clear ($n=32$, 15–53 rau)	0.366[a]							
Clear-conversational ($n=32$, −11–28 rau)	−0.358[a]	0.738[b]						
SPEAK[c] ($n=27$, 27–52)	0.205	0.234	0.081					
TOEFL[d] ($n=25$, 597–667)	−0.085	−0.250	−0.187	0.554				
AOA[e] ($n=31$, 5–18 years)	−0.063	−0.177	−0.132	−0.643	0.270			
LOA[f] ($n=31$, 6–17 years)	0.003	−0.139	−0.142	−0.071	−0.170	−0.436[a]		
LOI[g] ($n=32$, 0.02–2 years)	0.245	0.280	0.104	−0.421	0.094	0.008	−0.259	
Sentence production ($n=32$, 43–93 rau)	0.340[h]	0.449[b]	0.204	0.232	−0.229	−0.211	0.059	−0.061

[a] $p<0.05$.
[b] $p<0.01$.
[c] SPEAK = Speaking Proficiency English Assessment Kit (a test of spoken English).
[d] TOEFL = Test of English as a Foreign Language (mostly written).
[e] AOA = Age of Acquisition.
[f] LOA = Length of Acquisition.
[g] LOI = Length of immersion in an English speaking environment.
[h] $p<0.06$.

Finally, Fig. 3 shows individual subject data for the non-native listeners by native language background. The data are shown separately for those who responded to the female (top panel) and male (bottom panel) talkers. Within each plot, the data are presented in order of performance on the conversational speech materials (striped bars are in ascending order from left to right). The filled bars represent the amount of change in performance for the clear speech materials relative to the conversational speech materials. As shown in this figure, performance varied widely across listeners, even for listeners from the same native language background who responded to stimuli from the same talker. For example, performance of native speakers of Chinese covered a broad range in both baseline performance with conversational materials and in the magnitude of the clear speech benefit regardless of the talker, suggesting that native language background alone did not determine performance on the present test of sentence-in-noise perception.

IV. DISCUSSION

This study was designed to address three main issues: the magnitude of the clear speech effect across native and non-native listeners, the effect of increasing levels of noise on sentence perception by native and non-native listeners, and the factors that are related to sentence-in-noise perception abilities across individual non-native listeners. The findings of this study demonstrated that

(1) The group of non-native listeners exhibited a comparatively small clear speech benefit;
(2) The non-native listeners were not more adversely affected by increasing levels of noise than the native listeners;
(3) None of the demographic variables or test scores that we investigated was related to sentence-in-noise perception ability by the non-native listeners in this study. However, we found a significant negative correlation between the magnitude of the clear speech effect and conversational speech perception, and a significant positive correlation between sentence production ability and clear speech perception ability.

We discuss these finding in reverse order, ending with a discussion of the comparatively small clear speech benefit for the non-native listeners.

The lack of correlation between the various demographic variables and sentence-in-noise perception is somewhat surprising, given the results of earlier work that found significant relationships between some of these variables (e.g., age of acquisition and length of L2 study/exposure) and L2 speech perception performance (e.g., Flege, MacKay, and Meador, 1999). However, it is likely that the lack of a relationship between the sentence-in-noise perception test and any of the other factors that we investigated is due to the sample's uniformly limited amount of exposure to spoken English. Despite the fact that the subjects varied considerably in terms of age of English acquisition (5–18 years) and length of English study (6–17 years), none of these subjects had spent a significant amount of time in an English-speaking environment (none had spent more than 2 years, and all but two had spent under 1 year immersed in an English-speaking environment). In contrast, the non-native population studied by Flege and colleagues had long-term exposure to spoken English (several decades worth). Thus, while the non-native listeners in the present study showed a

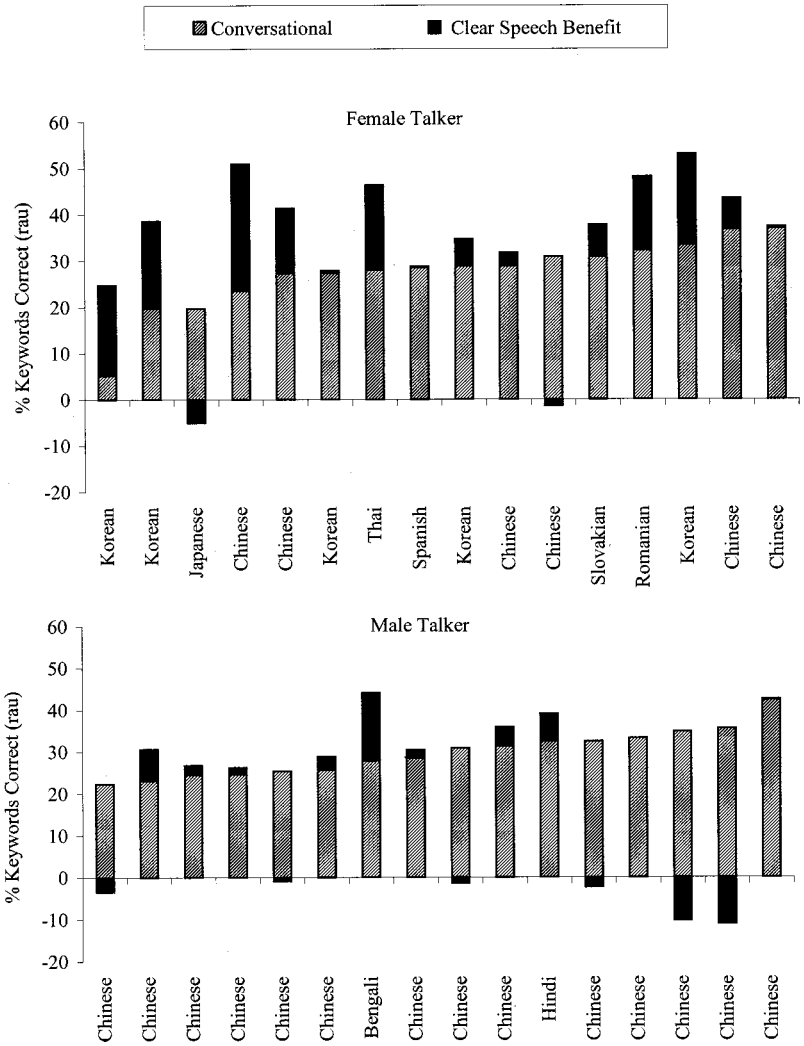


FIG. 3. Individual non-native subject's performance on the sentence-in-noise perception test broken down by native language background. Striped bars show performance with conversational speech stimuli; dark bars show the change in performance level from conversational to clear speech. Data for subjects who listened to the female and male talkers' productions are shown on the top and bottom panels, respectively.

considerable range of overall performance on the sentence-in-noise perception test (range = 39 rau; minimum = 15 rau; maximum = 54 rau), this variability could not be explained by any of the factors that we investigated. This result suggests that non-native listener spoken language processing in the absence of immersion in the target language depends on other factors, which still remain to be identified.

In contrast to the lack of correlation between demographic variables and the sentence-in-noise perception test, we found a significant negative correlation between the magnitude of the clear speech effect (i.e., the ability to take advantage of the talker-related clear speech modifications) and conversational speech perception ($r = -0.358$, $p < 0.05$). This relationship suggests that those non-native listeners with the most room for improvement are generally the ones who derive the most benefit from clear speech, and conversely that as sentence perception accuracy increases, the clear speech benefit decreases. This pattern of results seems to contradict the finding that the group of non-native listeners had a much smaller clear speech benefit than the native listeners, since the group of native listeners certainly had far higher overall speech perception accuracy than the group of non-native listeners. A possible resolution to this apparent paradox is that, whereas the native listeners derived most of their clear speech benefit from the language-specific, code enhancements of clear speech production (i.e., those that exaggerate the acoustic distance between contrasting categories and those that involve adherence to pronunciation norms), the non-native listeners derived most of their clear speech benefit from the signal enhancements of clear speech production (i.e., those that enhance the overall acoustic salience of the signal). Under this view, the negative correlation between conversational speech perception and the magnitude of the clear speech benefit for the non-native listeners is accounted for by supposing that the more proficient non-native listeners (as measured by the conversational speech perception accuracy) have shifted to a more native-like strategy of attempting to take advantage of the code enhancements of clear speech. While this strategy is ultimately efficient and effective, it requires extensive experience with the target language.

Initially, we hypothesized that the beneficial effects of the signal and code enhancements of clear speech production would be additive, that is, we hypothesized that all listeners would benefit from the signal enhancements and that listeners with experience with the spoken form of the target language would derive an additional benefit from the code enhancements. However, the pattern of results for the non-native listeners suggests that perhaps for normal-hearing non-native listeners, the beneficial effects of the signal en-

hancements are attenuated once the listener reaches a level of sufficient sentence perception proficiency where she or he begins to pay attention to the code enhancements of clear speech. This account for the apparently contradictory findings of (a) a negative conversational correlation between conversational speech perception and the magnitude of the clear speech benefit for the non-native listeners, and (b) a much smaller clear speech benefit for the group of non-native listeners than for the group of native listeners is highly speculative at this point and further investigation is required before any conclusive explanation can be offered.

The present study also found that the group of non-native listeners was not more adversely affected by noise than the group of native listeners; rather, the native listeners showed a greater difference between signal-to-noise ratio conditions than the non-native listeners for the male talker's productions. This result was surprising in view of previous findings to the contrary (Nábèlek and Donahue, 1984; Mayo, Florentine, and Buus, 1997; Meador, Flege, and McKay, 2000); however, it is possible that the non-native listeners in the present study exhibited a floor effect for the −8-dB signal-to-noise ratio condition. Future testing with inexperienced subjects such as those in the present study should involve more favorable signal-to-noise ratios so that overall performance can be elevated, thereby avoiding the possibility of a floor effect.

Furthermore, given that the native listeners in the present study showed different signal-to-noise ratio condition differences across the two talkers, it appears that some talkers may produce speech that is more resistant to noise than others. This talker-related factor is also seen when we compare the effect of speaking style across the two talkers (averaged across both signal-to-noise ratio conditions). For both the native and the non-native listeners, the two talkers had equivalent overall conversational speech intelligibility scores [native listeners: mean male versus female diff. = 2.8 rau, $t(30) = 0.86$, $p = 0.40$; non-native listeners: mean male versus female diff. = 2.3 rau, $t(30) = 0.97$, $p = 0.34$]. However, the female talker's clear speech intelligibility was significantly greater than the male's for the native listeners [mean male versus female diff. = 15.6 rau, $t(30) = 4.70$, $p < 0.0001$]. This difference approached significance for the non-native listeners [mean male versus female diff. = 6.21 rau, $t(30) = 1.99$, $p = 0.06$]. Furthermore, for the native listeners, the decline in intelligibility for the female's speech in the −8-dB signal-to-noise ratio condition (in the conversational speech condition) was smaller than the male's (male = −31 rau; female = −19 rau). Thus, although both talkers had comparable baseline intelligibility scores, the female talker was more effective at producing clear speech and her conversational speech was more resistant to increasing levels of noise.

In a companion study to the present study (Bradlow, Kraus, and Hayes, submitted), we report the results of a series of acoustic–phonetic measurements of the male's and female's productions that may be responsible for these differences. This analysis must be taken as preliminary because it was based on only two individual talkers; nevertheless, it revealed some notable differences between the clear speech productions of these two individual talkers. In particular, the acoustic measurements showed that for both talkers, conversational-to-clear speech modifications included a decreased speaking rate, longer and more frequent pauses, less alveolar stop flapping, more final stop releasing, greater consonant-to-vowel intensity ratios, a higher mean pitch, a wider pitch range, and an expanded vowel space. However, there were substantial differences between the degree to which each talker's clear speech exhibited these characteristics. In particular, the female decreased overall speaking rate and increased the frequency and duration of pauses far more than the male. The female also increased $F0$ mean and expanded her overall vowel space more than the male. However, the male avoided reduction processes, such as alveolar flapping and unreleased final stop consonants, and increased his pitch range more than the female. Given that the female's clear speech was significantly more intelligible than the male's (despite equivalent conversational speech intelligibility), this pattern of intertalker differences in clear speech production suggests that modifications to temporal characteristics, as well as increased articulatory effort and precision, may be particularly effective for enhancing intelligibility. In contrast, increased pitch range and elimination of ''reduction'' processes (such as alveolar flapping and final consonant releasing) may be less effective for enhancing intelligibility. It remains for further research with more talkers to determine whether these preliminary generalizations are valid.

The major purpose of the present study was to investigate whether non-native listeners would show a speech intelligibility benefit for naturally produced clear speech over conversational speech. This purpose was inspired by the finding of a robust clear speech effect for a wide range of listeners including hearing-impaired adults, elderly adults, and normal-hearing adults in degraded listening conditions. We were particularly interested in investigating the clear speech effect for non-native listeners, because the nature of their speech perception deficit is quite different from that of the other populations that have been the subjects of previous clear speech research. Whereas previous clear speech studies have focused on listeners whose knowledge of the target language is intact but their access to the speech signal is impeded, the present study focused on non-native listeners for whom the speech perception deficit results primarily from a lack of access to the underlying linguistic code. We reasoned that since clear speech production is essentially ''native-listener oriented,'' only those listeners who have experience with the phonology and phonetics of the target language would be able to take full advantage of the conversational-to-clear speech modifications. Nevertheless, we predicted that since clear speech production involves some features that are likely to increase intelligibility for all listeners, such as the slower speaking rate, wider dynamic pitch range, greater sound-pressure levels, more salient stop consonant releases, and greater obstruent rms intensity, non-native listeners would derive a significant, but relatively small, clear speech effect.

TABLE IV. Comparison of the clear speech effect size across various listener populations as reported in the literature (representative, but non-exhaustive).

Study	Materials	Listeners	Presentation conditions	Average clear speech effect
Payton *et al.*, 1994	nonsense sentences	2 HI[a] adults	OMCL freq-gain char. self-adjusted levels noise and reverb. as for NH 1 talker	26 rau
		10 NH[b] adults	3 reverberation conditions 3 noise conditions 3 reverb/noise combinations 1 talker	20 rau
Schum, 1996	meaningful sentences	60 HI adults	multitalker babble 10 young, 10 elderly talkers	20 rau
Picheny *et al.*, 1985	nonsense sentences	5 HI adults	2 freq.-gain characteristics 3 levels 3 talkers	17 rau
Helfer, 1997	nonsense sentences	39 NH adults	multitalker babble, 3 SNRs audio, visual, audiovisual	17 rau
Present study	meaningful sentences	32 NH adults	2 noise levels 2 talkers	16 rau
Bradlow *et al.* (submitted)	meaningful sentences	38 NL[c] children	2 noise levels 2 talkers	9 rau
		77 LI[d]children	2 noise levels 2 talkers	9 rau
Present study	meaningful sentences	32 NN[e]adults	2 noise levels 2 talkers	5 rau

[a]HI = hearing-impaired.
[b]NH = normal hearing.
[c]NL = normal hearing and not learning-impaired.
[d]LI = learning-impaired.
[e]NN = non-native.

The findings of the present study confirmed this prediction by demonstrating a far smaller clear speech effect for the non-native listener group than for the native listener group. Indeed, the average non-native listener clear speech effect of 5 rau was far smaller than the average native listener clear speech effect of 16 rau on this sentence-in-noise perception test. This native listener clear speech effect is comparable in magnitude to previously reported clear speech effects for both hearing-impaired listeners and normal listeners under degraded listening conditions, confirming that our test provided data that could be used for a valid comparison of the clear speech effect across several populations. Table IV presents such a comparison, where the data have been listed in descending order of the magnitude of the clear speech effect. Included in this table are the data from a companion study to the present study which used the same stimuli and experimental design as the present study to investigate the clear speech effect in normal and learning-impaired children (Bradlow, Kraus, and Hayes, submitted). While the data included in this table do not represent an exhaustive list of all clear speech perception studies, they are representative of the general range of the clear speech effect across various listener populations.

It is evident from the data in Table IV that for hearing-impaired listeners and normal listeners with degraded signals the clear speech effect is remarkably consistent across studies. Excluding the data from one study that found an unusually large effect of 26 rau for two hearing-impaired listeners (Payton *et al.*, 1994), the range of the effect is just 4 rau (16–20 rau). [If we include the data from the Payton *et al.* (1994) study, the range is 10 rau.] This consistency indicates that for these listeners the clear speech effect is robust and stable in magnitude across a variety of talker and listening conditions.

Of greater interest for the present study is the finding that the data in Table IV appear to fall into three groups: a group with a large clear speech effect (>15 rau), a group with an intermediate clear speech effect (9 rau), and a group with a small clear speech effect (5 rau). We propose that an important factor that distinguishes the three groups of listeners represented by these levels of the clear speech effect is the stage of target language development, and that this factor plays a major role in determining the likelihood that a listener will benefit from naturally produced clear speech. The listeners in the first group get the large >15-rau clear speech benefit because of their extensive experience with the target language. Thus, despite the fact that their access to the speech signal is impeded, they are able to take advantage of most, if not all, of the clear speech enhancements. The children in the second group (mean age = 10.6 and 10.4 years, s.d. = 1.8 and 2.2 years for the normal and learning-impaired children, respectively) represent a listener population in which the process of language acquisition is not complete. Thus, according to this interpretation, these children were able to derive some benefit from the clear speech enhancements, but this benefit was smaller than for the adults in the first group. (For more discussion of these data see Bradlow *et al.*, submitted.) The third group of subjects, the non-native

listeners (who had the least amount of experience with spoken English), showed only a very small clear speech effect. A prediction from this pattern of data is that the magnitude of the clear speech effect develops with age and exposure to the spoken language, and therefore that the individuals in the child and non-native listener groups will eventually attain the large clear speech benefit exhibited by the native listener adults.

A noteworthy limitation of the present data was the use of meaningful sentence materials in the sentence-in-noise perception test. Even though the non-native subjects in this study were quite proficient in English grammar (they all performed well enough on a standardized test of English reading and writing proficiency to be admitted into an American graduate school), it is very likely that they were less able than the native listeners to take advantage of the contextual information offered by the meaningful sentence materials. This important difference may have been particularly salient under the more difficult listening conditions, that is for the −8-dB signal-to-noise ratio and for the conversational speech materials. It is therefore possible that the difference in the magnitude of the clear speech benefit across the two groups of subjects in this study reflected a difference in their ability to take advantage of contextual information rather than a difference in their ability to take advantage of the acoustic–phonetic modifications of English clear speech. It remains for future research to address this important limitation.

We conclude by emphasizing the importance of broadening the range of listener populations that are tested on their ability to take advantage of naturally produced clear speech. It is very likely that different listener populations will find different enhancement features of clear speech more or less beneficial. Without a better understanding of how talker- and listener-related factors interact to influence overall speech intelligibility, we will remain limited in our ability to enhance speech intelligibility through the use of either naturally or synthetically produced clear speech over a wide range of communicative situations.

ACKNOWLEDGMENTS

We are grateful to Rebecca Schwarzlose and Juri Matsubara for data collection and processing, and to Xueqing Xu for programming assistance. We also thank two anonymous reviewers and Keith Kluender for extremely helpful comments on an earlier version of the manuscript. This research was supported by NIH-NIDCD Grant No. DC 03762.

APPENDIX: SENTENCE-IN-NOISE PERCEPTION TEST MATERIALS

List 1 (BKB-R List 7, keywords underlined)
(1) The children dropped the bag.
(2) The dog came back.
(3) The floor looked clean.
(4) She found her purse.
(5) The fruit is on the ground.
(6) Mother got a saucepan.
(7) They washed in cold water.
(8) The young people are dancing.
(9) The bus left early.
(10) They had two empty bottles.
(11) The ball is bouncing very high.
(12) Father forgot the bread.
(13) The girl has a picture book.
(14) The orange was very sweet.
(15) He is holding his nose.
(16) The new road is on the map.

List 2 (BKB-R List 8, keywords underlined)
(1) The boy forgot his book.
(2) A friend came for lunch.
(3) The match boxes are empty.
(4) He climbed his ladder.
(5) The family bought a house.
(6) The jug is on the shelf.
(7) The ball broke the window.
(8) They are shopping for cheese.
(9) The pond water is dirty.
(10) They heard a funny noise.
(11) The police are clearing the road.
(12) The bus stopped suddenly.
(13) She writes to her brother.
(14) The football player lost a shoe.
(15) The three girls are listening.
(16) The coat is on a chair.

List 3 (BKB-R List 9, keywords underlined)
(1) The book tells a story.
(2) The young boy left home.
(3) They are climbing the tree.
(4) She stood near her window.
(5) The table has three legs.
(6) A letter fell on the floor.
(7) The five men are working.
(8) He listened to his father.
(9) The shoes were very dirty.
(10) They went on a vacation.
(11) The baby broke his cup.
(12) The lady packed her bag.
(13) The dinner plate is hot.
(14) The train is moving fast.
(15) The child drank some milk.
(16) The car hit a wall.

List 4 (BKB-R List 10, keywords underlined)
(1) A dish towel is by the sink.
(2) The janitor used a broom.
(3) She looked in her mirror.
(4) The good boy is helping.
(5) They followed the path.
(6) The kitchen clock was wrong.
(7) The dog jumped on the chair.
(8) Someone is crossing the road.
(9) The mailman brought a letter.
(10) They are riding their bicycles.
(11) He broke his leg.
(12) The milk was by the front door.
(13) The shirts are hanging in the closet.
(14) The ground was very hard.
(15) The buckets hold water.
(16) The chicken laid some eggs.

Bamford, J., and Wilson, I. (**1979**). ''Methodological considerations and practical aspects of the BKB sentence lists,'' in *Speech-hearing Tests and the Spoken Language of Hearing-impaired Children*, edited by J. Bench and J. Bamford (Academic, London), pp. 148–187.
Bench, J., and Bamford, J., Editors (**1979**). *Speech-hearing Tests and the Spoken Language of Hearing-impaired Children* (Academic, London).
Best, C. T. (**1994**). ''The emergence of native-language phonological influences in infants: A perceptual assimilation model,'' in *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, edited by J. Goodman and H. C. Nusbaum (MIT Press, Cambridge, MA), pp. 167–224.
Best, C. T. (**1995**). ''A direct realist view of cross-language speech perception,'' in *Speech Perception and Linguistics Experience: Issues in Cross-language Research*, edited by W. Strange (York, Baltimore), pp. 171–204.
Bradlow, A. R., and Pisoni, D. B. (**1999**). ''Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors,'' J. Acoust. Soc. Am. **106**, 2074–2085.
Bradlow, A. R. (**in press**). ''Confluent talker- and listener-related forces in clear speech production,'' in *Papers in Laboratory Phonology VII*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin).
Bradlow, A. R., Kraus, N., and Hayes, E. (submitted). ''Speaking clearly for learning-impaired children: Sentence in noise perception,'' J. Speech Lang. Hear. Res. .
Cutler, A., and Butterfield, S. (**1990**). ''Durational cues to word boundaries in clear speech,'' Speech Commun. **9**, 485–495.
Ferguson, S. H., and Kewley-Port, D. (**in press**). ''Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am.
Flege, J. E. (**1992**). ''Speech learning in a second language,'' in *Phonological development: Models, Research, and Applications*, edited by C. Ferguson, L. Menn, and C. Stoel-Gammon (York, Timonium, MD), pp. 565–604.
Flege, J. E. (**1995**). ''Second language speech learning: Theory, findings, and problems,'' in *Speech Perception and Linguistics Experience: Issues in Cross-language Research*, edited by W. Strange (York, Baltimore), pp. 233–277.
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.
Hazan, V., and Simpson, A. (**2000**). ''The effect of cue-enhancement on consonant intelligibility in noise: Speaker and listener effects,'' Lang Speech **43**, 273–294.
Helfer, K. S. (**1997**). ''Auditory and auditory-visual perception of clear and conversational speech,'' J. Speech Lang. Hear. Res. **40**, 432–443.
Johnson, K., Flemming, E., and Wright, R. (**1993**). ''The hyperspace effect: Phonetic targets are hyperarticulated,'' Language **69**, 505–528.
Kuhl, P. K., and Iverson P. (**1995**). ''Linguistics experience and the perceptual magnet effect,'' in *Speech Perception and Linguistics Experience: Issues in Cross-language Research*, edited by W. Strange (York, Baltimore), pp. 121–154.
Lewellen, M. J., Goldinger, S. D., Pisoni, D. B., and Greene, B. G. (**1993**). ''Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision and semantic categorization,'' J. Exp. Psychol. Gen. **122**, 316–330.
Matthies, M., Perrier, P., Perkell, J. S., and Zandipour, M. (**2001**). ''Variation in anticipatory coarticulation with changes in clarity and rate,'' J. Speech Lang. Hear. Res. **44**, 340–353.
Mayo, L. H., Florentine, M., and Buus, S. (**1997**). ''Age of second-language acquisition and perception of speech in noise,'' J. Speech Hear. Res. **40**, 686–693.
Meador, D., Flege, J. E., and MacKay, I. R. (**2000**). ''Factors affecting the recognition of words in a second language,'' Bilingualism **3**, 55–67.
Moon, S.-J., and Lindblom, B. (**1994**). ''Interaction between duration, context, and speaking style in English stressed vowels,'' J. Acoust. Soc. Am. **96**, 40–55.
Nabelek, A. K., and Donahue, A. M. (**1984**). ''Perception of consonants in reverberation by native and non-native listeners,'' J. Acoust. Soc. Am. **75**, 632–634.
Ohala, J. J. (**1995**). ''Clear speech does not exaggerate phonemic contrast,'' in Proceedings of the 4th European Conference on Speech Communication and Technology, pp. 1323–1325.
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.
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.
Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1989**). ''Speaking clearly for the hard of hearing. III. An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech,'' J. Speech Hear. Res. **32**, 600–603.
Schum, D. J. (**1996**). ''Intelligibility of clear and conversational speech of young and elderly talkers,'' J. Am. Acad. Audiol **7**, 212–218.
Strange, W., Editor (**1995**). *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (York, Baltimore).
Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., and Nishi, K. (**2001**). ''Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners,'' J. Acoust. Soc. Am. **109**, 1691–1704.
Studebaker, G. A. (**1985**). ''A rationalized arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.
Uchanski, R. M., Choi, S. 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.
Uchanski, R. M. (**1988**) ''Spectral and temporal contributions to Speech Clarity for hearing impared listeners,'' Doctoral dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (unublished).
Uchanski, R. M. (**1992**) ''Segment durations in conversational and clear speech'' (unpublished).

The effect of channel interactions on speech recognition in cochlear implant subjects: Predictions from an acoustic model

Chandra S. Throckmorton and Leslie M. Collins
Department of Electrical and Computer Engineering, Duke University, Box 90291, Durham, North Carolina 27708-0291

(Received 24 November 2001; revised 4 April 2002; accepted 8 April 2002)

Acoustic models that produce speech signals with information content similar to that provided to cochlear implant users provide a mechanism by which to investigate the effect of various implant-specific processing or hardware parameters independent of other complicating factors. This study compares speech recognition of normal-hearing subjects listening through normal and impaired acoustic models of cochlear implant speech processors. The channel interactions that were simulated to impair the model were based on psychophysical data measured from cochlear implant subjects and include pitch reversals, indiscriminable electrodes, and forward masking effects. In general, spectral interactions degraded speech recognition more than temporal interactions. These effects were frequency dependent with spectral interactions that affect lower-frequency information causing the greatest decrease in speech recognition, and interactions that affect higher-frequency information having the least impact. The results of this study indicate that channel interactions, quantified psychophysically, affect speech recognition to different degrees. Investigation of the effects that channel interactions have on speech recognition may guide future research whose goal is compensating for psychophysically measured channel interactions in cochlear implant subjects.
 [DOI: 10.1121/1.1482073]

PACS numbers: 43.71.Ky, 43.71.Es, 43.71.An [CWT]

I. INTRODUCTION

Psychophysical experiments with cochlear implant subjects often reveal intersubject differences. For example, although sequential stimulation of electrodes from the base of the cochlea to the apex usually results in the perception of a monotonic change in pitch, pitch reversals have been observed by several investigators (e.g., Busby *et al.*, 1994; Collins *et al.*, 1997; Shannon, 1983a; Townshend *et al.*, 1987). As a second example, several studies have demonstrated that some subjects have one or more electrodes that are indiscriminable and that there may be a relationship between speech recognition and electrode discrimination (Henry *et al.*, 1997, 2000; Throckmorton and Collins, 1999; Zwolan *et al.*, 1997). Furthermore, numerous, detailed experiments have demonstrated that forward masking occurs even when masker and probe stimuli are not presented to the same electrode (Chatterjee, 1999; Chatterjee and Shannon, 1998; Lim *et al.*, 1989; Shannon, 1983b; Throckmorton and Collins, 1999). The results of these forward masking studies indicated that neural populations associated with different electrodes overlap and that the degree of overlap may vary across subjects. These examples illustrate three psychophysical measures that assess different aspects of channel interaction. Although pitch-ranking tasks do not technically measure channel interactions, since pitch is a characteristic of a single channel, the data from such tasks may be considered a measure of the underlying physiological anomalies that correspond to channel interactions. As such, for simplicity, references in this paper to channel interactions include references to anomalous pitch structures.

Given that channel interactions that can be quantified psychophysically occur in some implant subjects and may degrade speech recognition, methods by which to improve the individualized tuning of speech processors can be considered. However, investigating compensation methods for a particular channel interaction using cochlear implant subjects is complicated by several issues. First, different forms of interaction may be present in the same subject, and the effects may not be separable. Second, a new speech processor that does not immediately provide better speech recognition may adversely affect the quality of life of implanted subjects in the short term, even though performance may improve over time. Last, such studies often require more time than implanted subjects are able to commit to a study. Thus, the approach taken in this study is to independently compare the relative effects of a set of channel interactions on speech recognition performance using acoustic models that perform signal processing similar to that performed by two commercial speech processors. Future research on remediation procedures in cochlear implant subjects may then be focused on the channel interaction, psychophysically measured, that results in the greatest detriment to speech recognition.

Acoustic models of cochlear implants have the advantage of independent control of the experimental variables (e.g., Dorman *et al.*, 1997b; Fu *et al.*, 1998; Shannon *et al.*, 1998). Furthermore, the use of normal-hearing subjects provides several advantages, e.g., the subject pool can be trained quickly (Rosen *et al.*, 1999) and changes in speech-processing strategy do not adversely affect the quality of life of the subjects. Also, their speech recognition abilities are not dependent on demographic characteristics, as can be the case with cochlear implant subjects (e.g., Geier *et al.*, 1999; Rubinstein *et al.*, 1999; Shipp and Nedzelski, 1995; van Dijk *et al.*, 1999). Some trends in cochlear implant psychophysi-

cal and speech recognition performance that have been successfully predicted by acoustic models include the saturation of pitch as a function of pulse rate at approximately 300 pps (Blamey *et al.*, 1984), the superior performance of subjects with a simple spectral mapping algorithm as opposed to a feature-extraction algorithm (Tong *et al.*, 1990), the pattern of vowel and consonant recognition for nonlinear mapping compression functions (Fu and Shannon, 1998), and the sensitivity of subjects to spectral shifting (Dorman *et al.*, 1997a; Fu and Shannon, 1999).

Studies comparing data from cochlear implant subjects and data from normal-hearing subjects measured via acoustic models of cochlear implants have also revealed some limitations regarding the validity of conclusions about cochlear implant data based on acoustic modeling results. Although the acoustic model created by Blamey *et al.* (1984) successfully predicted the saturation of pitch as a function of pulse rate at approximately 300 pps, the change in pitch was much greater for cochlear implant subjects than for the normal-hearing subjects. Similarly, in the study of nonlinear amplitude mapping by Fu and Shannon (1998), the pattern of percent-correct scores for vowel and consonant recognition was similar for cochlear implant subjects and normal-hearing subjects; however, the peak scores were typically about 10 percent higher for the normal-hearing subjects. Furthermore, cochlear implant subjects were found to be more sensitive to signal-to-noise ratio (SNR) than normal-hearing subjects using the acoustic model (Fu *et al.*, 1998).

Comparing responses due to acoustic stimulation to those elicited by electrical stimulation must be approached with caution as the neural responses of these two types of stimulation inherently differ. Although phase locking occurs for both acoustic and electrical stimulation, the discharge patterns differ considerably since electrical stimulation produces a highly synchronous response (e.g., Hartmann *et al.*, 1984; Kiang and Moxon, 1972). On the other hand, the higher-order processing conducted by the nervous system can be assumed to remain the same. Thus, when acoustic models are used to study cochlear implants, results should be interpreted in terms of the trends that are predicted rather than as a quantitative estimate of cochlear implant subject performance.

A final, important consideration when using an acoustic model to investigate trends in cochlear implant performance is the ability of the normal-hearing subjects to adapt to the distorted stimuli. To date, the ability to adapt to adverse conditions has primarily been investigated in the context of spectrally shifted speech that represents the variation of insertion depths possible for implant subjects. Dorman *et al.* (1997a), using an acoustic model which varied the shift between analysis frequencies and the stimulation frequencies to which they are mapped, found a significant decrease in speech recognition when such a mismatch occurs. These results have been replicated in several other studies (Fu and Shannon, 1999; Rosen *et al.*, 1999; Shannon *et al.*, 1998); however, Rosen *et al.* (1999) further demonstrated that the performance of normal-hearing subjects could improve with training and experience.

A similar result was observed for cochlear implant subjects in a study by Fu and Shannon (1999). The analysis frequency bands were shifted while the stimulation frequencies (i.e., the electrode positions) to which they were mapped remained the same. Although electrode arrays tend to be located more basally than the analysis filters mapped to them in implant subjects, the highest performance for the subjects did not result from basally shifted analysis filters. The highest performance instead resulted from analysis filters most closely aligned with their clinically assigned analysis filters, i.e., subjects performed best with the systems to which they had grown accustomed (Fu and Shannon, 1999). This suggests that cochlear implant subjects can also, to some degree, overcome spectral degradation (in this case, mismatch) with training and experience. Thus, measurement of the effects of channel interactions through acoustic models, e.g., detriment to speech recognition due to spectral shift in frequency mapping, may be more applicable to cochlear implant subjects without years of experience with their device.

This last study raises the question of whether attempting to correct psychophysically measured channel interactions is worthwhile, since cochlear implant subjects can overcome these defects at least to some degree through adaptation. Trying to correct interactions that result in serious degradations of speech recognition is probably worthwhile during the initial tuning of the speech processor for new cochlear implant users, as this might reduce the amount of time necessary for new users to reach their peak performance. Also, although cochlear implant subjects adapt with experience, assessing whether they reach the best performance possible is difficult. For these reasons, understanding the relative effects of channel interactions is worthwhile. However, given that subjects do adapt, any attempts at compensation for channel interactions in experienced cochlear implant users may initially lead to worse performance. This decrease in performance was observed in a study by Collins *et al.* (1997) when remapping filters to electrodes to reflect pitch-ranking data. The necessity of training implant subjects with new algorithms prior to assessing performance should be considered.

With these caveats in mind, a study was designed to assess the relative impact of different psychophysically measured channel interactions. A standard acoustic model (termed the unaltered model) was designed to serve as the control. The standard model is similar to that used in other modeling studies (e.g., Dorman *et al.*, 1997b; Fu *et al.*, 1998; Shannon *et al.*, 1998). Channel interactions are modeled by changing the structure of the standard model. Three types of channel interactions are modeled in this study: nontonotopic pitch order, indiscriminable electrodes, and variable forward-masking patterns. The speech recognition scores measured with these models were compared to determine which interaction produced the greatest degradation in speech recognition. Based on the results from the initial data collection, two additional models were designed: pitch gap models and a modified set of forward masking models. The speech recognition results from these models were compared to the results from the pitch order and original forward masking models, respectively.

TABLE I. A list of the model acronyms.

Acronym	Model description
6/20F	Twenty bandpass filter speech-processing algorithm for which data from only six filters are presented in each analysis cycle.
8F	Eight bandpass filter speech-processing algorithm.
ED case I	Electrode discrimination model for which each pair of band-pass filters is replaced by a single filter encompassing the passbands of both original filters.
ED case II	Electrode discrimination model for which each pair of carrier signal frequencies is identical.
FBI	Forward masking model, intermediate case, whose masked thresholds have been adjusted by a constant to produce an average proportion of masked stimuli approximately equal to the average produced by the FMB model.
FBW	Forward masking model, worst case, whose masked thresholds have been adjusted by a constant to produce an average proportion of masked stimuli approximately equal to the average produced by the FMB model.
FIW	Forward masking model, worst case, whose masked thresholds have been adjusted by a constant to produce an average proportion of masked stimuli approximately equal to the average produced by the FMI model.
FMB	Forward masking model, best case: masker only affects probes of the same frequency.
FMI	Forward masking model, intermediate case: masker affects all probes with greater masked thresholds for those probes closer in frequency to the masker.
FMW	Forward masking model, worst case: masker affects all probes equally.
PGH	Pitch gap, high frequency.
PGL	Pitch gap, low frequency
PGM	Pitch gap, mid frequency.
PRH	Pitch reversal, high frequency.
PRL	Pitch reversal, low frequency.
PRM	Pitch reversal, midfrequency.

II. MODELS

In this section, we describe the standard, or unaltered model, as well as the models used to simulate various channel interactions. All of the models described below were implemented in MATLAB®. Acronyms have been assigned to each model due to the large number of models and due to the complexity of the models that makes assigning simple and yet descriptive names to the models difficult. These acronyms are defined with the model descriptions below, and the acronyms are listed in Table I for reference purposes.

A. Unaltered model

Speech tokens were first prefiltered for spectral equalization (first-order high-pass Butterworth filter with cutoff frequency of 1 kHz). The speech tokens were then passed through a low-pass sixth-order Butterworth filter to prevent aliasing by removing frequencies greater than half the sampling frequency. The specific speech tokens that were used are described in detail in the Methods section.

Two different algorithms were used to process the speech and were designed to be similar to the processing performed by either the CIS (e.g., Kessler, 1999; Wilson *et al.*, 1991) or SPEAK (e.g., Whitford *et al.*, 1995) processing strategy. Both of these algorithms perform a mapping of the information in different frequency bands to location in the cochlea. The difference between them is mainly a tradeoff between resolution and information content. The CIS algorithm presents all of the available information at the cost of spectral resolution, while the SPEAK algorithm has higher spectral resolution but presents only the highest energy information. Another consideration, not modeled in this study, is that CIS allows more rapid stimulation of the electrodes resulting in greater temporal resolution. These two algorithms were considered because they mimic speech-processing techniques currently in use.

The prefiltered speech was processed either by a bank of eight bandpass filters with logarithmically spaced central frequencies (8F algorithm) that spanned a frequency range of 150–6450 Hz, or by a bank of 20 bandpass filters with linearly spaced central frequencies below 1600 Hz and logarithmically spaced central frequencies above 1600 Hz, spanning a range of 250–10 823 Hz. The 8F algorithm presented acoustic information from all eight filters. The second algorithm only presented acoustic information from the six filters with the greatest energy for each window of speech, and is thus referred to as the 6/20F algorithm. All bandpass filters were sixth-order Chebyshev type I filters. The envelopes of the bandpass filters were extracted using full-wave rectification and low-pass filtering (eighth-order Chebyshev type I low-pass filter). Acoustic presentation consisted of a summation of the carrier signals, the amplitudes of which were derived from the root-mean-square (rms) energy of the envelopes computed every 2 ms, similar to the method described by Dorman *et al.* (1997b). The carrier signals consisted of sinusoids, similar to the design implemented by Dorman *et al.* (1997b). The authors assumed an ideal situation in which the simulated ''electrode'' was placed equidistant from the cutoff frequencies of the corresponding bandpass filters. Thus, the carrier signal frequencies were computed by first converting the cutoff frequencies of the bandpass filters to distances along the cochlea using the Greenwood function (Greenwood, 1990), averaging these distances, and then converting the average distance back to frequency using the Greenwood function.

B. Pitch-order models

Pitch reversals that have been observed psychophysically (e.g., Collins and Throckmorton, 2000; Collins *et al.*, 1997; Shannon, 1983a; Townshend *et al.*, 1987) are modeled acoustically by changing the frequency of the carrier signals, and these changes are the only difference between these models and the unaltered model. Based on psychophysical data, 25 percent of the carrier frequencies were assumed to be affected by the reversal (Collins and Throckmorton, 2000; Collins *et al.*, 1997). Five carrier signal frequencies were changed to create each pitch reversal for the 6/20F algorithm, while only two frequencies were changed for the 8F algorithm (see Fig. 1). Three types of reversals were tested: low frequency, midfrequency, and high frequency, referred to as PRL (pitch reversal low-frequency), PRM, and PRH, respectively, as the effects from reversals might be frequency dependent.

Psychophysical measurements from cochlear implant subjects do not provide information regarding the exact fre-

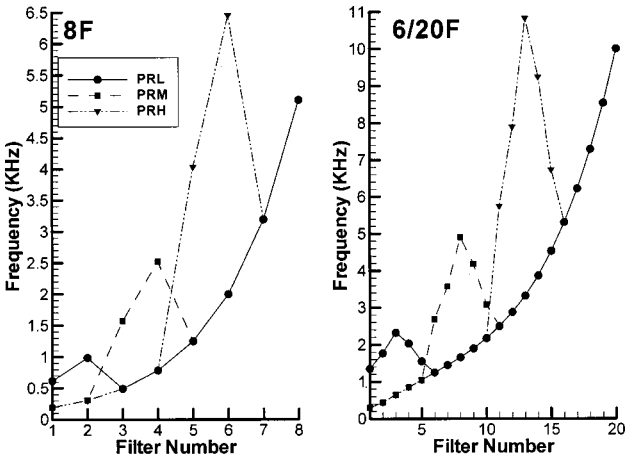


FIG. 1. These plots depict the assignments of carrier signal frequencies to analysis filters for the three pitch-reversal models. The left panel shows the mapping of the carrier signal frequencies for the 8F algorithm and the right panel shows the mapping for the 6/20F algorithm. Note that the scale for the two plots is different since the two algorithms have different frequency ranges. The carrier signal frequencies for two of the bandpass filters are remapped by each reversal for the 8F algorithm, while five carrier signal frequencies are remapped for the 6/20F algorithm.

quencies stimulated, since pitch-ranking measurements determine only relative pitch order and not absolute magnitude of the frequency shifts. Thus, the choice of the new frequencies is an inherently arbitrary process. Without data on which to base the choice of the reversal frequencies, these frequencies were chosen based on other criteria. The reversal frequencies were chosen to be equidistant and between unaltered frequencies to maximize discrimination between frequencies, and changes to the pitch range were minimized.

C. Pitch-gap models

Pitch reversals could negatively affect speech recognition via several mechanisms. First, the shifting of temporal information from one frequency band into a higher frequency band may confound the temporal information already contained within the higher frequency band. Second, since the temporal information is shifted, no temporal information is presented in the original frequency band. For example, if the temporal information for a vowel's first formant is presented in the second formant's frequency band, the shifted temporal information would compete with and possibly confuse the temporal information carried by the second formant. Furthermore, no temporal information would be presented in the first formant's frequency band, a frequency band that would normally contain temporal information.

To further investigate the factors underlying the effects of pitch reversals on speech recognition, three models were designed for which the carrier signals affected by the reversals were discarded rather than assigned new frequencies. In effect, a dead zone is created in which the spectral and temporal information of a speech signal contained within a particular frequency band is discarded while the information contained within the other frequency bands remains unchanged. After discarding the carrier signals, the 8F algorithm presents only six tones per window of speech. The 6/20F algorithm continues to presents six tones per window; however, the total number of filters is reduced from 20 to 15. The frequency band for which the carrier signals (and thereby the temporal information) are discarded is hereafter referred to as a pitch gap. Three models, whose pitch gaps correspond to the reversals in the pitch-order models, were designed: the pitch-gap low-frequency model (PGL), pitch-gap midfrequency model (PGM), and the pitch-gap high-frequency model (PGH). Thus, the three pitch-gap models were designed to be compared to the pitch-order models rather than being based upon psychophysical measurements in cochlear implant subjects. Since the confounding temporal information was not presented in the pitch-gap models, the hypothesis was that speech recognition measured for the pitch-gap models should be greater than that measured for the pitch-reversal models.

D. Electrode discrimination

Indiscriminable electrodes result in a situation in which two (or more) identically perceived frequency bands carry different temporal information. Thus, although having indiscriminable electrodes is a sign of reduced spectral resolution, the amount of temporal information remains the same. The question arises as to whether it is more detrimental to combine the temporal information for identically perceived frequency bands by creating one bandpass filter that encompasses the frequencies of the indiscriminable electrodes or to allow redundant frequency bands to present different temporal information. Thus, two models were created for this study: one in which the number of analysis bandpass filters is reduced and their passbands are broadened (ED case I); and one in which the number of passbands of the analysis filters remains the same, but some of the carrier signals have the same presentation frequency (ED case II). In order to weight the effects of the electrode indiscriminability evenly across frequency, every other filter was considered indiscriminable, i.e., filters 1 and 2 are indiscriminable, filters 3 and 4 are indiscriminable, etc. This design criterion was chosen over frequency-dependent models in light of the already large scale of the study.

Zwolan *et al.* (1997) measured speech recognition for cochlear implant subjects with their clinical mapping of filters to electrodes, and with the removal of indiscriminable electrodes. The method for removing these indiscriminable electrodes is similar to the ED case I, where the passband of the analysis filter is expanded to encompass the passbands of the filters mapped to the indiscriminable electrodes, and this new, larger bandwidth filter is then mapped to one of the indiscriminable electrodes, and the others are not stimulated. Thus, if electrodes 1 and 2 are indiscriminable, then the bandwidth of filter 1 is expanded to encompass the passband of filter 2 as well, and then only electrode 1 is stimulated. Using this method to correct for electrode discrimination, Zwolan *et al.* (1997) measured significant increases in speech recognition for five out of nine subjects on at least one speech-recognition measure. Thus, the results observed in Zwolan *et al.* (1997) were the motivation for implementing the ED case I model.

Considering every other filter to be indiscriminable, the ED case I model reduces spectral resolution by 50% by com-

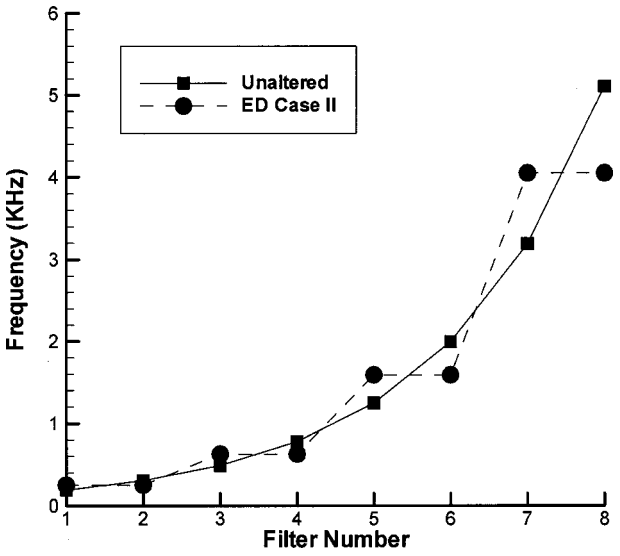


FIG. 2. This plot shows the carrier signal frequencies for the unaltered and ED case II models for the 8F algorithm. Each pair of filters is mapped to the same carrier frequency, and that frequency is centrally located between the mappings of the unaltered model. The carrier frequencies used for each filter pair are the same as the carrier frequencies used for each expanded bandpass filter in the ED case I model.

bining each pair of bandpass filters into one filter. Thus, the 8F algorithm would have four bandpass filters, each with a unique carrier frequency, and the 6/20F algorithm would choose six out of ten unique carrier signals. The carrier frequencies were determined for the broadened filters using the same method described for the unaltered model.

The second case (ED case II) investigates redundant spectral presentation by leaving the bandpass filters the same as in the unaltered model, but changing the carrier frequencies such that each pair of successive carrier frequencies is the same. The frequencies assigned to each carrier signal are the same as those determined for the ED case I model, e.g., the first two carrier signals would have the same frequency as the first carrier signal in the ED case I model. Figure 2 demonstrates the difference between the unaltered model and the ED case II model, plotting carrier signal frequency versus filter number. For the ED case II model, each pair of filters is mapped to the same frequency and these frequencies are centrally located between the mappings of the unaltered model. Thus, the spectral information has been reduced by 50%; however, the temporal information remains the same as that provided by the unaltered model.

E. Forward masking

Forward masking data from implant subjects have shown that, as in normal-hearing subjects, one stimulus can mask subsequent stimuli. The forward masking models are based on estimating whether a stimulus is detectable (i.e., not masked) given a prior stimulation pattern, and when masked, removing that stimulus from the stimulation pattern by setting the amplitude to zero. Similar to the method described in Throckmorton and Collins (1999), masking patterns were used to remove individual pulses from the stimulation pattern. Each window was considered to provide a set of potential maskers, and subsequent windows were considered to be sets of probe stimuli. Thus, for the 8F algorithm, each window would consist of eight maskers and subsequent windows would contain eight probe stimuli. Similarly, for the 6/20F algorithm, each window would consist of six maskers with subsequent windows containing sets of six probe stimuli. The effect of each masker on all the probe stimuli was considered by comparing the amplitudes of the probe stimuli to the masked thresholds defined by the masking patterns. If the probe amplitudes were lower than the masked thresholds, then the probe stimuli were considered to be masked and removed from the stimulation pattern for a particular speech token.

The difficulty in modeling forward masking lies in the design of the masking patterns. To use this method for removing stimuli, a masking pattern must exist for each combination of masker and probe, as a function of variables such as masker amplitude and probe delay (the time between masker offset and probe onset). Without a complete set of psychophysical data such as this for a set of cochlear implant subjects, some simplifying assumptions were necessary. First, the peak in the masking pattern was assumed to occur when the masker and probe are presented to the same electrode. The masking pattern was also assumed to decrease linearly from the peak masking level with increasing electrode separation and to be symmetric around the peak. Furthermore, the pattern was not considered to be a function of masker frequency. Although these assumptions are not necessarily mirrored in the psychophysical data, the simplified design enabled the investigation of the relative effects of forward masking on speech recognition.

Once the static pattern described above was designed, rules were developed to modify the pattern for dynamic factors such as masker amplitude and probe delay. In a study by Chatterjee and Shannon (1998), increasing the masker amplitude resulted in increases in masking levels in a roughly additive fashion. Based on this result, changes in masker amplitude were modeled as additive changes in the masking pattern.

Chatterjee (1999) further investigated, for identical probe and masker stimuli, the relationship between masked threshold and probe delay. The masked threshold was found to decrease exponentially as the probe delay increased. Chatterjee (1999) reported that the exponential decay of the masked threshold appeared to have both a rapidly decaying component as well as a slowly decaying component, with the slowly decaying component dependent on the masker duration. Considering the short duration of the masker in this study (2 ms), the slower-decaying component was considered to have no impact on the masking patterns. The time constant of the rapidly decaying component used in this study was chosen to be 3.75 ms, roughly the middle of the range of time constants reported by Chatterjee (1998) when fitting curves to the data from three cochlear implant subjects. Thus, for the first set of probe stimuli, the probe delay was zero, and the masking pattern was unchanged; however, each subsequent window of probe stimuli were compared to the masking pattern calculated by

$$\mathrm{MT}_N = (\mathrm{MT}_O)e^{-2(w^P - w^M - 1)/3.75},$$

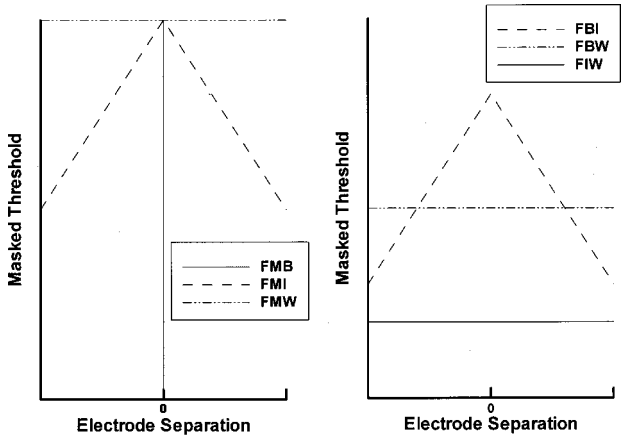


FIG. 3. An example of the differences between the masking patterns for the forward masking models. Note: these patterns are not drawn to scale. In the left panel are representations of the three models that differ in the amount of spectral interactions. For the best case, FMB, masking only occurs when the masker and probe are identical, and for the worst case, FMW, all probes are equally affected by the masker. In the right panel are representations of the additional three forward masking models that were designed. Constants were subtracted from the masked thresholds of the FMI and FMW models in order to achieve approximately equal proportions of masked stimuli with the FMB and FMI models.

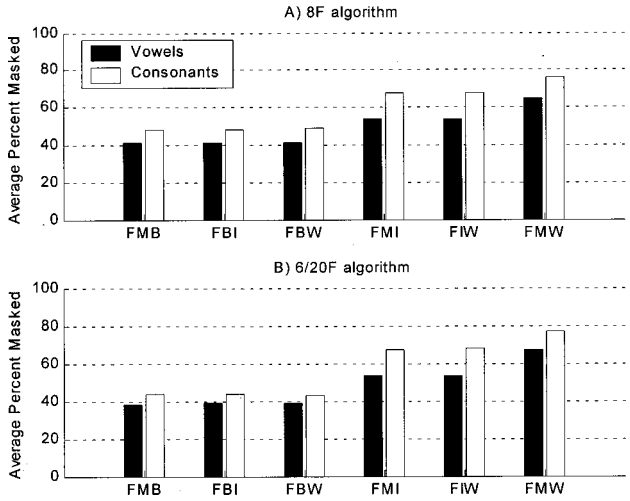


FIG. 4. A bar chart showing the average proportion of masked stimuli for vowel and consonant tokens for (A) the 8F algorithm, and (B) the 6/20F algorithm. The FMB, FMI, and FMW models represent a best, intermediate, and worst case for forward masking patterns, respectively. The FBI and FBW models are the FMI and FMW models with the masked thresholds adjusted to produce average proportions of masked stimuli approximately equal to the average for the FMB model. The FIW model is the FMW model with its masked thresholds adjusted to produce an average approximately equal to that produced by the FMI model.

where MT_N is the set of masked thresholds, MT_O is the original set of masked thresholds, and w^P and w^M are the window numbers of the masker and probe stimuli, respectively.

1. Spectral interactions

Three masking patterns were chosen to evaluate different levels of spectral interaction: the forward masking best case (FMB), in which maskers only affected probes of the same frequency; the worst case, in which maskers affected probes of all frequencies equally (FMW); and the intermediate case, in which maskers affected probes of all frequencies but effects were greater for those frequencies closer to the masker's frequency (FMI). The left panel of Fig. 3 shows an example of the masking patterns for these three cases by plotting masked thresholds as a function of electrode separation. The amount of masking caused by the FMI model was determined by slowly increasing the amount of masking from the FMB model until the difference between speech processed by the FMB model and speech processed by the FMI model were just detectable. The average proportion of stimuli removed is shown for vowel and consonant recognition for the three models in Fig. 4. Thus, these three models all have the same peak masked threshold; however, the models ranked lowest to highest in terms of spectral interactions would be the FMB, FMI, and FMW models. Similarly, the models would be ranked in the same order for the average proportion of masked stimuli.

2. Comparison of spectral and temporal interactions

Lower masked thresholds in a masking pattern can indicate fewer temporal interactions, and a smaller set of probes affected by the masker can indicate fewer spectral interactions. In order to compare temporal and spectral interactions, three models were designed such that the amount of temporal and spectral interactions varied but the average proportion of masked stimuli remained constant. The first two models were generated by subtracting a constant from all masked thresholds for the FMI and FMW models such that the average proportion of masked stimuli approximately matched that produced by the FMB model, and these new models were termed the FBI and FBW models, respectively. Similarly, the FMW model was adjusted again such that it produced approximately the same average proportion of masked stimuli as the FMI model, resulting in the FIW model. Examples of the masking patterns for these three models are shown in the right panel of Fig. 3. The average proportions of masked stimuli for all of the forward masking models are shown in Fig. 4. The FMB, FBI, and FBW models all have approximately the same average proportion of masked stimuli; however, the models ranked in terms of peak masked threshold from lowest to highest would be FBW, FBI, and FMB models, and ranked in terms of the spectral interactions from lowest to highest would be FMB, FBI, and FBW. Similarly, the FMI model has higher temporal and lower spectral interactions than the FIW model.

III. METHODS

A. Subjects and equipment

Twelve subjects were recruited for this study; however, one subject was dropped from the study due to high thresholds at some frequencies that were detected during an audiogram prior to the experiment. Four female and seven male subjects completed the initial data collection, and one female and three male subjects from the same subject pool completed the second data collection. The subjects were recruited from the undergraduate and graduate student population. Three of the subjects had previous experience participating

in psychoacoustic experiments, and one subject also had familiarity with the speech-recognition tasks used. None of the subjects had been exposed to the speech processed by the acoustic models prior to this study.

Before data collection, subjects' thresholds for tones with frequencies of 250, 500, 1000, 2000, 4000, and 8000 Hz were measured to verify that no significant hearing loss was present. Tones for threshold measurements were presented via equipment purchased through Tucker-Davis Technologies (TDT). A 2 down/1 up 2-interval forced choice (2IFC) threshold task was designed using software provided by TDT (1997). The threshold task as well as the model-processed speech-recognition tasks were performed on an IBM compatible computer in a soundproof booth (IAC, Bronx, NY) with the stimuli presented acoustically to the subjects through Sony MDR-V600 dynamic stereo headphones. Subjects responded to the tasks by typing their response on the keyboard or by using the mouse to click a button. No feedback was provided.

B. Stimuli

Four different speech recognition tasks were used to judge the effects of different models. Vowel recognition was tested using a nine-choice (had, hawed, head, heard, heed, hid, hood, hud, who'd) closed-set medial vowel test in which each token was presented five times in random order (Tyler *et al.*, 1986). Consonant recognition was tested using a 14-choice (b, d, f, g, j, k, m, n, p, s, sh, t, v, z) closed-set medial consonant test with consonants presented in the form ''aCa'' presented five times randomly (Tyler *et al.*, 1986). Sentence recognition was tested using the CID Everyday Sentence Lists 1–5 (Davis and Silverman, 1978), and monosyllabic word and phoneme recognition was measured using the NU #6 Monosyllabic Words Lists 1–4 (Tillman and Carhart, 1966). The speech tokens were all sampled from the Cochlear Corporation/The University of Iowa Revised Cochlear Implant Test Battery CD (Cochlear, 1995) at a sampling frequency of 22 050 Hz.

The amount of data that could be gathered from a single subject using the sentence, word, and phoneme recognition was limited. While an almost infinite number of tests for vowel and consonant recognition could be generated by randomly creating a new list that presents each token five times, the other tasks could not use repeated tokens as familiarity with the tokens would bias the performance results. Thus, the data that could be gathered using these tokens for any subject were limited to one presentation of each test provided on the CD, while vowel and consonant recognition tests could be generated independently of the CD.

The speech tokens processed by the models were not processed in real time. All speech was processed prior to presentation and saved as WAV files that were then presented to the subjects as needed. Subject responses were recorded using an interface custom designed in LABVIEW™. Subjects typed their responses for the sentence, word, and consonant recognition tasks. Since the consonant recognition task is closed set, a list of the possible responses was displayed by the monitor. For the vowel recognition task, nine buttons, each labeled with a token, were displayed on the monitor and subjects responded by clicking the correct button.

C. Data collection

1. Training

The goal of this study was to investigate the level to which different channel interactions affected speech recognition. Thus, removing competing factors such as familiarity with the tasks or the processed speech was necessary. Therefore, the first session for the subjects consisted of training. These data were not used in the data analysis. The subjects were trained on vowel and consonant recognition tokens processed by the unaltered model for both algorithms. The subjects were tested with 5–7 repetitions of each combination of recognition task and algorithm. Subjects were considered trained when their performance plateaued, rather than whether their scores exceeded a preset threshold.

2. Initial data collection

Vowel and consonant recognition was tested in all the subjects under nine modeled conditions: the unaltered model, PRL, PRM, PRH, ED case I, ED case II, FMB, FMI, and FMW models. All subjects were tested under these conditions for both the 8F and 6/20F algorithms. Two to four repetitions were carried out for each recognition task and each algorithm for all the models. Thus, at a minimum, subjects participated in 72 tests (2 recognition tasks $\times 2$ algorithms $\times 9$ models $\times 2$ repetitions). The order of these tests was randomized to eliminate training effects.

In addition to vowel and consonant recognition, sentence, word, and phoneme recognition was tested. However, due to the restraints discussed previously, not all models could be tested for all recognition tasks for each subject, as this would have required 18 unique sentence tests and 18 unique word lists (2 algorithms $\times 9$ models), when only 5 tests and 4 lists were available. Therefore, each subject was tested on five sentence tests and four word lists for nine, randomly chosen combinations of speech processing algorithm and model, with no combination repeated for a single subject. In this way, 2–4 measures of sentence, word, and phoneme recognition could be recorded for each combination across the subject pool. The nine sentence and word tests were interspersed throughout the vowel and consonant recognition tasks in order to lessen the fatigue that subjects experience after repeated vowel and consonant tasks.

3. Follow-up data collection

The analysis of the data collected during the initial experiment led to further research hypotheses. In order to investigate these hypotheses, six additional models (pitch-gap and forward masking comparison of interactions models) were designed, and the speech recognition of four of the better-performing subjects (S2, S5, S6, S10) was measured with the additional models. Training was reverified at the beginning of the study. Subjects were presented with 1–2 repetitions of speech processed by the unaltered models for each of the four combinations of recognition task and algorithm. The training data were then graded to verify that sub-

jects were still performing at their peak level. Further training was provided if subjects misidentified more than five tokens from either test.

Once training had been verified, the subjects were tested under seven test conditions: the unaltered, PGL, PGM, PGH, FBI, FBW, and FIW models. Testing was performed identically to the initial data collection. Only vowel and consonant recognition were measured, as the four subjects had been exposed to all available sentence and word lists. Each combination of recognition task, algorithm, and model was tested 2–3 times for each subject, resulting in a minimum of 56 tests.

IV. RESULTS

A. Training

As demonstrated by Rosen (1999), the performance for normal-hearing subjects listening to speech through an acoustic model of cochlear implant speech processing improves with training. Subjects in this study were also observed to improve with experience, often achieving scores close to the maximum possible during training. On average, subjects' vowel scores increased from the initial measurements by 35% while their consonant scores increased by 7%–10%. For the 8F algorithm, the average across subjects of the maximum scores achieved during training was 92% correct (standard deviation = 12.2) for vowel recognition and 94% correct (standard deviation = 5.8) for consonant recognition. For the 6/20F algorithm, the average maximum scores were 97% correct (standard deviation = 6.5) for vowel recognition and 96% correct (standard deviation = 4.8) for consonant recognition. The high standard deviations are due to a few subjects plateauing to a different (substantially lower) level than other subjects; however, on average, the subjects achieved scores close to the maximum possible score during training.

B. Comparison of channel interactions

The speech-recognition data for vowel and consonant recognition were pooled across subjects and placed into confusion matrices according to the combination of algorithm, model, and recognition task. From the confusion matrices, the percent-correct scores were determined and are plotted in Fig. 5. The "*" indicates scores that are significantly different ($p < 0.05$) from the unaltered model scores. Statistical significance was determined using confidence intervals based on the assumption that the subject responses could be considered to be binomially distributed (Thornton and Raffin, 1978). Specifically, the method put forth by Thornton and Raffin (1978) for calculating the confidence intervals involves first transforming the scores using an arcsine transform (Freeman and Tukey, 1950) to achieve uniform variance. Using the transformed data, the variance between the difference in scores can be calculated, and the probability of a difference occurring can be estimated.

Almost all of the models produced scores that were significantly lower than the unaltered model scores, although the differences were sometimes small. The PRH, ED case II, and FMB models did not produce significantly lower scores

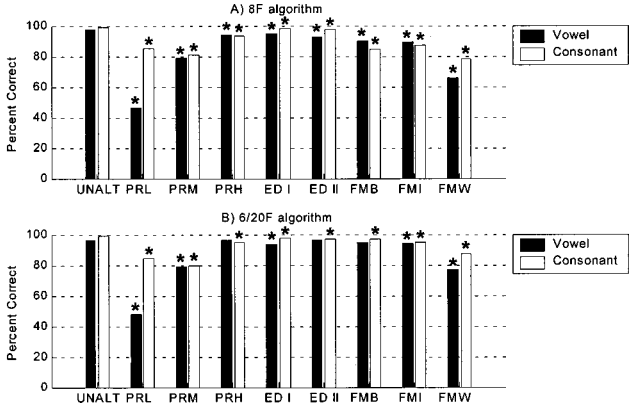


FIG. 5. The vowel and consonant percent-correct scores from data pooled across subjects for each model used in the initial data collection for (A) the 8F algorithm and (B) the 6/20F algorithm. The "*" indicates scores statistically different from the scores obtained with the unaltered model ($p < 0.05$).

than the unaltered model for vowel recognition under the 6/20F algorithm. Since a limit exists on the maximum achievable score, a difficulty arises in judging whether the difference between these scores is not significant. By using a nonlinear transform of the data, such as the arcsine transform described above (Thornton and Raffin, 1978), the ceiling effects on judgments of statistical significance may be partially alleviated. Furthermore, since the goal of the study is to search for those models producing the greatest decrease in speech-recognition scores, recognizing that these scores are much higher than those from other models is sufficient.

For vowel recognition, the PRL, PRM, and FMW models produce the worst scores for both algorithms. For consonant recognition, these three models again have detrimental effects; however, for the 8F algorithm, the other two forward masking models result in similar degradations. These results are similar to those measured for the sentence, monosyllabic word, and phoneme recognition scores (see Fig. 6). The scores plotted in Fig. 6 are the average of results from only 2–4 subjects, as discussed in the methods section, and thus, the results should be considered with caution. However, in general, the results agree with the pattern of results observed in the vowel and consonant recognition task results.

C. Comparison of pitch models

If pitch reversals are hypothesized to result in two perceptual problems, confounding temporal information and pitch gaps, a system that consists only of pitch gaps might be hypothesized to result in better speech-recognition scores. This hypothesis is supported by the data collected using the 6/20F algorithm. Figure 7 shows the difference in scores for the pitch-order models and the pitch-gap models. The scores for the pitch-order models for this comparison were determined by using only the data for the four subjects tested under the pitch-gap models so that the results from the two model types would be from identical subject groups. A significant increase ($p < 0.05$) in scores occurs between the PRL and PGL models as well as the PRM and PGM models. The lack of a statistically significant increase in scores for the PGH model may be due to a ceiling effect since the scores

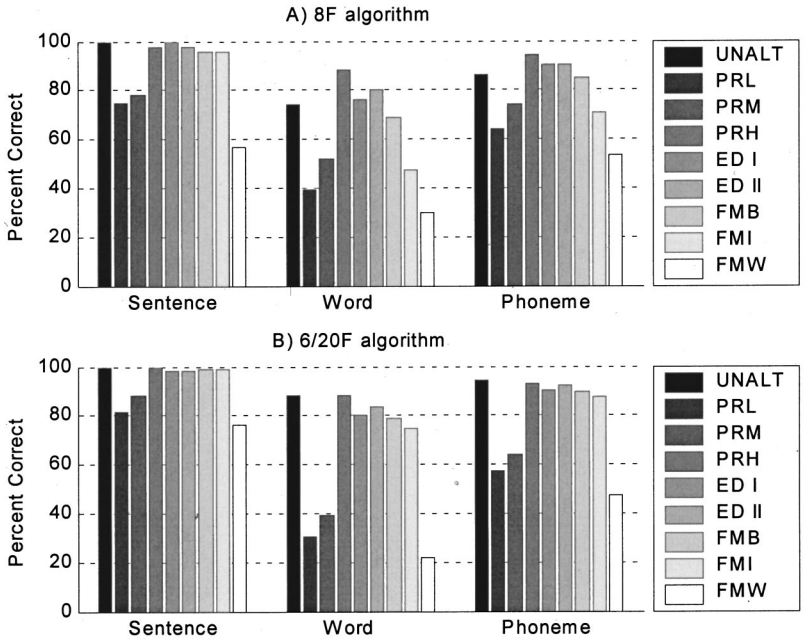


FIG. 6. The sentence, monosyllabic word, and phoneme percent-correct scores from data pooled across subjects for each model used in the initial data collection for (A) the 8F algorithm and (B) the 6/20F algorithm. Note that all phonemes must be correctly detected for the word responses to be correct; therefore, phoneme percent-correct scores must be greater than or equal to word percent-correct scores.

from the PRH model are already close to the maximum. Thus, for the PRL and PRM models, the confounding temporal information appears to have a detrimental effect in addition to the effect caused by a gap in the pitch order.

The results for the 8F algorithm, however, imply that the relationship between the pitch-gap models and the pitch-order models may be dependent on characteristics of the speech-processing algorithms such as number of filters or filter bandwidths. Although the 6/20F algorithm scores suggest that under most circumstances it is better to discard reversal stimulations than to leave them unaltered, the 8F algorithm scores suggest that this may not always be the case (see Fig. 7). Although the PGL model significantly outperforms the PRL model, the PRM and PGM models have scores that are not significantly different for vowel recognition, and the PGH model produces scores significantly *worse* than the PRH model.

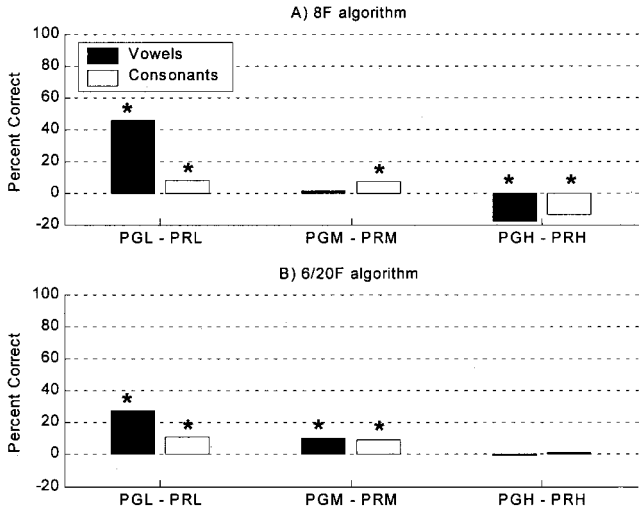


FIG. 7. The difference between the percent-correct scores for the pitch-order and pitch-gap models, e.g., the PGL model scored approximately 25% correct higher than the PRL model for the 6/20F algorithm. The "*" indicates statistical difference ($p<0.05$) between the scores of the two models.

D. Comparison of forward masking models

Figure 8 shows the difference in scores between the FMB and FMI masking models and the FBI, FBW, and FIW masking models. Similar to the comparison between pitch models, the scores for all the forward masking models were determined from the data from four subjects so that the results from the two model types would be from identical subject groups. The percent-correct score from the FMB model is subtracted from the percent-correct scores from the FBI and FBW models since these models have approximately the same average proportion of masked stimuli. Similarly, the score from the FMI model is subtracted from the score from the FIW model. For the 8F algorithm, consonant recognition is consistently better ($p<0.05$) for the model with the lower

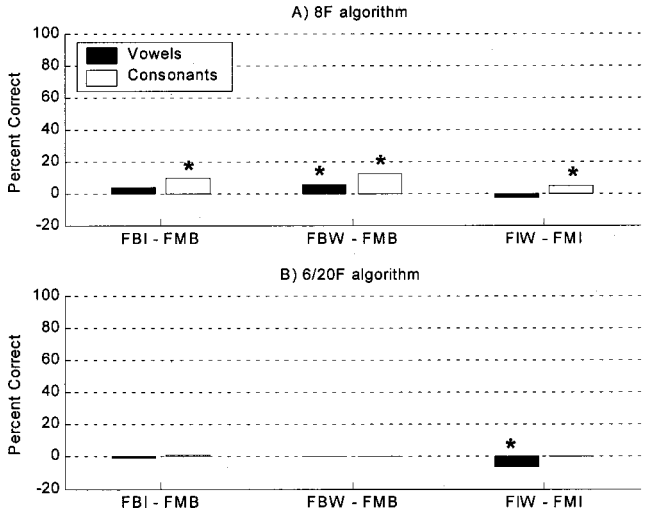


FIG. 8. The difference between the scores of forward masking models with approximately the same proportion of masked stimuli but different levels of temporal and spectral interactions. The score of the model with the greater masked thresholds and lesser spectral spread was subtracted from the score of the model with the lower masked thresholds and greater spectral spread. The "*" indicates statistical difference ($p<0.05$) between the scores of the two models.

temporal interactions. Similarly, for vowel recognition, the one statistically higher score ($p<0.05$) occurred for the model with the lower temporal interactions. Thus, although these groups of models have approximately the same average proportion of masked stimuli, decreasing the temporal interactions improves the scores significantly. Furthermore, the results suggest that the more the masked thresholds are lowered and the spectral spread increased, the more improvement in the scores. Note that the difference between the FBI and FMB models is approximately 9%, while the difference between the FBW and FMB models is approximately 12%. The difference between the FBI and FBW model scores is statistically significant ($p<0.05$). For the 6/20F algorithm, statistical significance between scores occurred in only one case: the vowel recognition scores for the FMI and FIW models. Opposite to the results from the 8F algorithm, the higher the temporal interactions and lower the spectral interactions produced significantly better scores ($p<0.05$). Since significance occurred in only one case, however, conclusions are difficult to infer about this algorithm.

V. DISCUSSION

Comparison of the forward masking models provided some interesting insight regarding the effects of different masking patterns on speech recognition. For vowel recognition, statistical significance between the scores from models with approximately the same average proportion of masked stimuli occurred for only one comparison per speech-processing algorithm. The implication is that vowel recognition depends more on the average proportion of masked stimuli than on the relative masked thresholds and spectral spread. For the 6/20F algorithm, consonant recognition also seemed to be independent of the level of temporal and spectral interactions. However, with consonant recognition for the 8F algorithm, the data from the models with greater spectral spread tended to significantly outperform the data from models with greater masking levels. Thus, fewer temporal interactions with the same average proportion of masked stimuli appear to have a significant advantage for consonant recognition for the 8F algorithm.

The comparison of pitch models also provided some insight into the effects of pitch reversals. Although data collected using the 6/20F algorithm support the hypothesis that pitch reversals result in two phenomena, a gap in pitch representation and confounding temporal information, that potentially affect speech recognition, data from the 8F algorithm suggest that the relationship is more complicated. The relative error caused by pitch gaps and pitch reversals may depend on such factors as the frequencies affected, the number of carrier signals affected, the extent of the mismatch, etc.

The conflict between the results for the two algorithms may be due to the definitions of the reversals that differ in the frequencies that were affected (see Table II). The data from the pitch gap models agree to some degree with electrode discrimination data gathered from cochlear implant subjects. In an electrode discrimination study of cochlear implant subjects using the SPEAK algorithm, Henry *et al.* (2000) found a correlation between electrode discrimination and speech recognition only for the frequency bands below 2680 Hz. All of the reversals for the 8F algorithm affect frequencies below 2680 Hz, and all of the pitch gap models produce scores statistically lower than the unaltered model ($p<0.05$). For the 6/20F algorithm, only the low- and mid-frequency reversals have frequencies completely below 2680 Hz, and only these pitch-gap models produce scores statistically lower than the unaltered model ($p<0.05$).

TABLE II. The frequency ranges (in Hz) affected by the pitch-gap models.

	PGL	PGM	PGH
8F	150–384	384–984	984–2519
6/20F	250–1150	1150–2323	2323–4903

Henry *et al.* (2000) also found the highest correlation between electrode discrimination and speech recognition for the frequency band 1768–2680 Hz. For the 8F algorithm, the PGH model encompasses this frequency range and the PGH model produced the lowest scores of the three pitch-gap models. For the 6/20F algorithm, the PGM model encompasses this frequency range. The PGM model did produce the lowest scores for this algorithm for consonant recognition but not for vowel recognition. Thus, further investigation may be warranted to provide greater insight into these results and their relation to electrode discrimination data. Although psychophysically based models address the goals of this study, future work may need to rely on models not based on measurements in cochlear implant subjects in order to determine the exact mechanisms responsible for the degradation of speech recognition. For example, further insight might be gained from models that, while not based on psychophysical measurements in cochlear implant subjects, take a systematic approach to the removal of frequency bands. This might provide information about the relative importance of different frequency bands for speech recognition.

The results from Henry *et al.* (2000) may also explain in part the results from the ED case II models. The models were designed to avoid bias in any particular frequency band; however, the results from Henry *et al.* (2000) suggest that frequency band is an important factor affecting the relationship between speech recognition and electrode discrimination. Furthermore, an equal reduction in spectral resolution due to indiscriminable electrodes across the array is unlike what is actually measured in cochlear implant subjects. These factors may explain in part why the results from the ED case II models are only slight, albeit typically significant, decreases in speech recognition. Future investigation into this area might involve models designed in a manner similar to the pitch-gap models in which different frequency bands are systematically removed.

In the case of the ED case I model and 8F algorithm, the lower scores observed by Dorman and Loizou (1997) and Dorman *et al.* (1997b) for a four-channel model may be explained by the difference in training that subjects received in those studies and this study. Dorman *et al.* (1997b) presented all tokens twice prior to commencing testing; however, subjects were not familiarized with stimuli to the extent of subjects in this study. In addition to a difference in training, the

model in this study differed from that of Dorman *et al.* (1997b) in that 2-ms-window durations were used for discretizing the envelope rather than 4 ms. Thus, the model in this study had twice the amount of temporal information than that used by Dorman *et al.* (1997b). The increased amount of temporal information and the active training may be responsible for the difference between the results of this study and those of the previous study.

Overall, spectral anomalies appear to have had the greatest potential for degrading speech recognition. Furthermore, this conclusion appears to hold for both the 8F algorithm, which presents all of the available temporal information at lower spectral resolution, and the 6/20F algorithm, in which only a subset of temporal information is presented at a higher spectral resolution. The similarity in the results across the two processing algorithms suggest that the robustness of these two algorithms appears to be dependent on the fundamental similarity between these algorithms in the method of presenting temporal and spectral information rather than on the algorithm-specific methods of selecting the temporal information to present. Thus, the greatest degradation occurs when the fundamental processing of the algorithms does not perform as designed.

Although spectral anomalies appear to have the greatest potential for degrading speech, the effect appears to be frequency dependent. Speech recognition is less robust to variations in the lower frequencies. This result is similar to that from the electrode discrimination study by Henry *et al.* (2000) in cochlear implant subjects that suggests that the importance of discrimination is frequency dependent. Furthermore, the results seen in this study agree with modeling studies of mismatched filters and carrier signal frequencies that indicate that speech recognition is sensitive to spectral anomalies (Dorman *et al.*, 1997a; Fu and Shannon, 1999; Shannon *et al.*, 1998).

Forward masking did decrease speech recognition in some cases; however, the effects never reached the level of degradation in closed-set recognition that was observed for some of the pitch reversal models. However, in this study the forward masking model focused on short-term masking effects—the rapidly decaying component observed by Chatterjee (1999). Further masking effects might be seen if the slowly decaying component was implemented for cases in which the masker is longer in duration due to sustained energy in specific frequency bands across several windows, e.g., as might be observed for vowel stimulation. Future work investigating the effects of longer maskers may reveal more insight into the degradation in speech recognition caused by forward masking. However, in terms of comparing forward masking effects to those of other channel interactions, the models produce a proportion of masked stimuli that may be much higher than is actually experienced by cochlear implant subjects. Based on their data, Throckmorton and Collins (1999) suggest that the average proportion of masked stimuli may be more on the order of 5%–19%. Thus, although for open-set recognition the FMW model did decrease speech recognition to a similar level as the low-frequency pitch reversals, levels of masked stimuli closer to the levels experienced by cochlear implant subjects would probably reduce the negative effects of the masking models.

In this study, the two algorithms considered were based on algorithms already in use by cochlear implant subjects. As such, these results suggest that remediation research working within the framework of these algorithms should focus on spectral anomalies, especially for the lower-frequency region. However, it is possible that simple remediation strategies within the framework of these algorithms may not be possible. Future work might consider whether speech-processing strategies that differ fundamentally from these two algorithms also show a lack of robustness to low-frequency spectral anomalies when compared to the other types of channel interactions used in this study. However, in terms of the original goal of this work to focus remediation research under current speech-processing algorithms, the results presented here suggest that future remediation work in implant subjects may benefit from focusing on compensation of low-frequency spectral anomalies.

ACKNOWLEDGMENTS

We express our thanks and appreciation to the subjects who gave us their time and effort. We also thank Dr. Chris Turner, Dr. Philip Loizou, and an anonymous reviewer for their comments regarding this study. This research was supported by NSF under Grant No. NSF-BES-00-85370.

Blamey, P. J., Dowell, R. C., Tong, Y. C., and Clark, G. M. (**1984**). ''An acoustic model of a multiple-channel cochlear implant,'' J. Acoust. Soc. Am. **76**, 97–103.

Busby, P. A., Whitford, L. A., Blamey, P. J., Richardson, L. M., and Clark, G. M. (**1994**). ''Pitch perception for different modes of stimulation using the Cochlear multiple-electrode prosthesis,'' J. Acoust. Soc. Am. **95**, 2658–2669.

Chatterjee, M. (**1999**). ''Temporal mechanisms underlying recovery from forward masking in multielectrode-implant listeners,'' J. Acoust. Soc. Am. **105**, 1853–1863.

Chatterjee, M., and Shannon, R. V. (**1998**). ''Forward masking excitation patterns in multielectrode electrical stimulation,'' J. Acoust. Soc. Am. **103**, 2565–2572.

Cochlear Corporation/The University of Iowa Revised Cochlear Implant Test Battery (**1995**). Englewood, CO.

Collins, L. M., and Throckmorton, C. S. (**2000**). ''Investigating perceptual features of electrode stimulation via a multidimensional scaling paradigm,'' J. Acoust. Soc. Am. **108**, 2353–2365.

Collins, L. M., Zwolan, T. A., and Wakefield, G. H. (**1997**). ''Comparison of electrode discriminability, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear impant subjects,'' J. Acoust. Soc. Am. **101**, 440–455.

Davis, H., and Silverman, S. R. (**1978**). Hearing and Deafness (Holt, Rinehart, and Winston, New York).

Dorman, M. F., and Loizou, P. C. (**1997**). ''Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants,'' Am. J. Otol. **18**, S113–S114.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997a**). ''Simulating the effect of cochlear-implant electrode insertion depth on speech understanding,'' J. Acoust. Soc. Am. **102**, 2993–2996.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997b**). ''Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs,'' J. Acoust. Soc. Am. **102**, 2403–2411.

Freeman, M. F., and Tukey, J. W. (**1950**). ''Transformations related to the angular and the square root,'' Ann. Math. Stat. **21**, 607–611.

Fu, Q.-J., and Shannon, R. V. (**1998**). ''Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners,'' J. Acoust. Soc. Am. **104**, 2570–2577.

Fu, Q.-J., and Shannon, R. V. (**1999**). ''Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing,'' J. Acoust. Soc. Am. **105**, 1889–1900.

Fu, Q.-J., Shannon, R. V., and Wang, X. (**1998**). ''Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,'' J. Acoust. Soc. Am. **104**, 3586–3596.

Geier, L., Fisher, L., Barker, M., and Opie, J. (**1999**). ''The effect of long-term deafness on speech recognition in postlingually deafened adult Clarion cochlear implant users,'' Ann. Otol. Rhinol. Laryngol. Suppl. **177**, 80–83.

Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Hartmann, R., Topp, G., and Klinke, R. (**1984**). ''Discharge patterns of cat primary auditory fibers with electrical stimulation of the cochlea,'' Hear. Res. **13**, 47–62.

Henry, B. A., McKay, C. M., McDermott, H. J., and Clark, G. M. (**1997**). ''The relationship between speech information perceived by cochlear implantees in different spectral regions and electrode discrimination,'' 1997 Conference on Implantable Auditory Prostheses, Asilomar Conference Center, Pacific Grove, California.

Henry, B. A., McKay, C. M., McDermott, H. J., and Clark, G. M. (**2000**). ''The relationship between speech perception and electrode discrimination in cochlear implantees,'' J. Acoust. Soc. Am. **108**, 1269–1280.

Kessler, D. K. (**1999**). ''The Clarion multistrategy cochlear implant,'' Ann. Otol. Rhinol. Laryngol. Suppl. **177**, 8–16.

Kiang, N. Y.-S., and Moxon, E. C. (**1972**). ''Physiological considerations in artificial stimulation of the inner ear,'' Ann. Otol. Rhinol. Laryngol. **81**, 714–730.

Lim, H. H., Tong, Y. C., and Clark, G. M. (**1989**). ''Forward masking patterns produced by intracochlear electrical stimulation of one and two electrode pairs in the human cochlea,'' J. Acoust. Soc. Am. **86**, 971–980.

Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). ''Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants,'' J. Acoust. Soc. Am. **106**, 3629–3636.

Rubinstein, J. T., Parkinson, W. S., Tyler, R. S., and Gantz, B. J. (**1999**). ''Residual speech recognition and cochlear implant performance effects of implantation criteria,'' Am. J. Otol. **20**, 445–452.

Shannon, R. V. (**1983a**). ''Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics,'' Hear. Res. **11**, 157–189.

Shannon, R. V. (**1983b**). ''Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction,'' Hear. Res. **12**, 1–16.

Shannon, R. V., Zeng, F.-G., and Wygonski, J. (**1998**). ''Speech recognition with altered spectral distribution of envelope cues,'' J. Acoust. Soc. Am. **104**, 2467–2476.

Shipp, D. B., and Nedzelski, J. M. (**1995**). ''Prognostic indicators of speech recognition performance in adult cochlear implant users: A prospective analysis,'' Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 194–196.

TDT (**1997**). PSYCHOSIG. Gainesville, FL, Tucker-Davis Technologies, Inc.

Thornton, A. R., and Raffin, M. J. M. (**1978**). ''Speech discrimination scores modeled as a binomial variable,'' J. Speech Hear. Res. **21**, 507–518.

Throckmorton, C. S., and Collins, L. M. (**1999**). ''Investigation of the effects of temporal and spatial interactions on speech-recognition skills in cochlear-implant subjects,'' J. Acoust. Soc. Am. **105**, 861–873.

Tillman, T. W., and Carhart, T. (**1966**). ''An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University Auditory Test No. 6,'' USAF School of Aerospace Medicine (Brooks Air Force Base, Texas).

Tong, Y. C., Harrison, J. M., Huigen, J., and Clark, G. M. (**1990**). ''Comparison of two speech processing schemes using normal-hearing subjects,'' Acta Oto-Laryngol., Suppl. **469**, 135–139.

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

Tyler, R. S., Preece, J. P., and Tye-Murray, N. (**1986**). The Iowa Phoneme and Sentence Tests, Iowa City, The University of Iowa, Department of Otolaryngology.

van Dijk, J. E., van Olphen, A. F., Langereis, M. C., Mens, L. H. M., Brokx, J. P. L., and Smoorenburg, G. F. (**1999**). ''Predictors of cochlear implant performance,'' Audiology **38**, 109–116.

Whitford, L. A., Seligman, P. M., Everingham, C. E., Antognelli, T., Skok, M. C., Hollow, R. D., Plant, K. L., Gerin, E. S., Staller, S. J., McDermott, H. J., Gibson, W. R., and Clark, G. M. (**1995**). ''Evaluation of the Nucleus Spectra 22 Processor and new speech processing strategy (SPEAK) in postlingually deafened adults,'' Acta Oto-Laryngol. **115**, 629–637.

Wilson, B. S., Lawson, D. T., Finley, C. C., and Wolford, R. D. (**1991**). ''Coding strategies for multichannel cochlear prostheses,'' Am. J. Otol. Suppl. **12**, 56–61.

Zwolan, T. A., Collins, L. M., and Wakefield, G. H. (**1997**). ''Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects,'' J. Acoust. Soc. Am. **102**, 3673–3685.

Effect of bone cortical thickness on velocity measurements using ultrasonic axial transmission: A 2D simulation study

Emmanuel Bossy,[a] Maryline Talmant, and Pascal Laugier
Laboratoire d'Imagerie Paramétrique, UMR CNRS 7623, Université Paris VI, 75006 Paris, France

(Received 12 October 2001; accepted for publication 29 March 2002)

In recent years, quantitative ultrasound (QUS) has played an increasing role in the assessment of bone status. The axial transmission technique allows to investigate skeletal sites such as the cortical layer of long bones (radius, tibia), inadequate to through-transmission techniques. Nevertheless, the type of propagation involved along bone specimens has not been clearly elucidated. Axial transmission is investigated here by means of two-dimensional simulations at 1 MHz. We focus our interest on the apparent speed of sound (SOS) of the first arriving signal (FAS). Its dependence on the thickness of the plate is discussed and compared to previous work. Different time criteria are used to derive the apparent SOS of the FAS as a function of source–receiver distance. Frequency-wave number analysis is performed in order to understand the type of propagation involved. For thick plates (thickness$>\lambda_{bone}$, longitudinal wavelength in bone), and for a limited range of source–receiver distances, the FAS corresponds to the lateral wave. Its velocity equals the longitudinal bulk velocity of the bone. For plate thickness less than λ_{bone}, some plate modes contribute to the FAS, and the apparent SOS decreases with the thickness in a way that depends on both the time criterion and on the source–receiver distance. The FAS corresponds neither to the lateral wave nor to a single plate mode. For very thin plates (thickness$<\lambda_{bone}/4$), the apparent SOS tends towards the velocity of the lowest order symmetrical vibration mode (S_0 Lamb mode).
 [DOI: 10.1121/1.1480836]

PACS numbers: 43.80.Ev, 43.80.Vj, 43.80.Jz, 43.80.Qf [FD]

I. INTRODUCTION

In recent years, quantitative ultrasound (QUS) has demonstrated its ability to detect bone fragility as efficiently as x-ray absorptiometry techniques,[1] which measure bone mineral density. Clinical devices that measure the speed of sound (SOS) and/or broadband ultrasound attenuation (BUA) are now widely used. To derive those parameters, most of the commercially available systems use through-transmission techniques, for which two transducers, one emitter and one receiver, are set on opposite sides of the skeletal site to be tested. Such techniques are well suited to peripheral skeletal sites such as the heel or the finger phalanges, but are inadequate to assess long bones such as the radius or the tibia. Indeed, the use of transmission techniques on such sites is made rather awkward, due to complex bone anatomy. Another approach, the so-called ''axial transmission'' has been developed to study propagation along the cortical shell of long bones, and was first applied in the late 1950s to study fracture healing.[2] In axial transmission technique, both the emitter and the receiver are in contact with the skin, placed on a same side of the skeletal site, and the first arriving signal (FAS) at the receiver is used to derive an apparent SOS. We point out now that the term SOS suggests that the velocity of a well-identified wave is measured. For reasons that will be clarified later in this work, we will prefer the term apparent SOS. The FAS has been presented[3–5] as a wave emitted from the surface at the longitudinal critical angle. Camus *et al.*[6] have documented this type of propagation in the case of an interface between two semi-infinite media (fluid/solid), which corresponds to the lateral wave, or *P*-head wave, well-known in seismology. This wave propagates along the interface between the fluid and the solid with the longitudinal bulk velocity of the solid. Nevertheless, the type of propagation occurring along bone specimens, of finite thickness and curved geometry, has not been clearly elucidated yet. Recently, it has been shown experimentally[7] that the apparent SOS of the FAS is dependent on the thickness of the studied sample. As the cortical thickness changes with aging and osteoporosis,[8] and has been shown to be a risk factor for fracture, it is critical to elucidate the relationship between the apparent SOS and the cortical thickness.

This paper is an extension of a previous work reported by Camus *et al.*[6] to the case of bone plates immersed in a fluid. Using two-dimensional (2D) numerical simulations of 1 MHz ultrasound propagation, we focus on the apparent SOS for bone plates in water, in order to study the relationship between apparent SOS and cortical thickness. Attention is paid to the accuracy of the measurement and to the impact of the signal processing techniques on measured values. We present a simple geometrical interpretation that predicts the existence of a range of source–receiver distances, depending on the plate thickness and on the way the signal is detected, for which the apparent SOS equals that measured in the case of a single interface between water and bone. Outside this range, the dependence of the apparent SOS on the plate thickness is discussed and compared to published works.

[a]Electronic mail: emmanuel.bossy@lip.bhdc.jussieu.fr

II. THEORETICAL BACKGROUND

A. The lateral wave

1. Introduction

We first consider the emission of a broadband spherical wave by a source S located in water near a plane interface separating water and bone. As the longitudinal bulk velocity of bone (approximately 4000 m/s) is greater than that of water (1490 m/s), such a configuration leads to the propagation of the so-called lateral wave. This type of propagation has been documented in detail by Camus *et al.*[6] For sake of clarity, we will briefly recall here the basic principles, by giving an illustration in the simplest case of two homogenous fluid half-spaces separated by a plane interface. The associated theoretical developments can be found in a work by Brekhovskikh and Godin.[9] The bulk wave velocity denoted c_w in the upper medium (water, which contains the source) is assumed to be lower than that of the bulk wave velocity in the lower medium, denoted c_l. The critical angle θ_c is defined by

$$\sin(\theta_c)=\frac{c_w}{c_l}. \tag{1}$$

When the spherical incident wave impacts the interface, it first gives rise to a reflected wave front and a refracted wave front [Fig. 1(a)]. As c_l is greater than c_w, the divergence of the refracted wave front is greater than that of the incident wave. As long as the angle of incidence θ_I is smaller than θ_c, the geometrical point I that connects the direct and the reflected wave front also connects them to the refracted wave front. With increasing time, this point I moves along the interface with the velocity defined by

$$v(I)=\frac{c_w}{\sin(\theta_I)} \tag{2}$$

with $v(I)>c_l$. When θ_I becomes greater than the critical angle θ_c, the velocity $v(I)$ of the geometrical point I becomes lower than velocity c_l. The refracted wave front is then no longer connected to point I and point J outdistances point I along the interface [Fig. 1(b)]. The theory then predicts the existence of a linear wave front which connects the refracted wave front (point J) to the reflected wave front (point K). This wave is called the lateral wave, or head wave. It propagates in water with velocity c_l, perpendicularly to the wave front, and makes an angle θ_c with the interface. It thus has an apparent velocity along the interface given by

$$v_{\text{lateral}}=\frac{c_w}{\sin(\theta_c)}=c_l. \tag{3}$$

Therefore, by measuring the time-of-flight (TOF) of the signal received at different positions aligned parallel to the interface, one can measure directly the bulk velocity c_l in the lower medium, provided the lateral wave is the first wave to reach the receivers. Before detailing the conditions required to observe the lateral wave as the first wave, let us predict what occurs in the case of a plane interface between a fluid and a solid. If the longitudinal bulk velocity of the solid c_l, given by

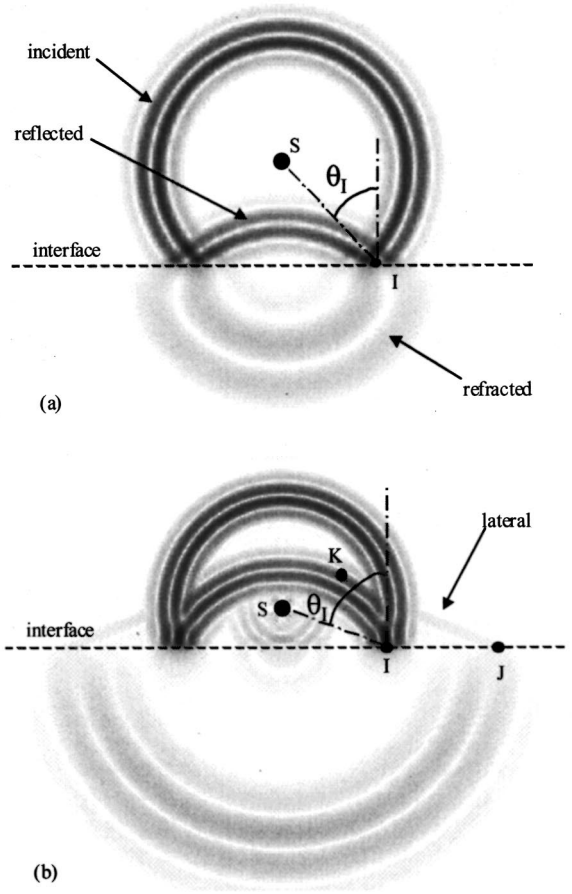


FIG. 1. Snapshots of the wave propagation, in the fluid/fluid case. (a) Before the refracted wave front has disconnected from the incident wave front. (b) After the refracted wave front has disconnected from the incident wave front.

$$c_l=\sqrt{\frac{\lambda+2\mu}{\rho}} \tag{4}$$

(λ, μ, and ρ being the Lamé coefficients and density of the solid), is still greater than the bulk velocity in the fluid, what has been said previously remains applicable. The lateral wave in the fluid/solid case connects the longitudinal refracted wave front to the reflected wave front, and propagates along the surface with the longitudinal bulk velocity in the solid c_l. The differences come from the many types of waves that can be generated in solid. In addition to the longitudinal refracted wave front, the incident spherical wave front will also generate a transverse refracted wave front, as well as other kinds of waves that may propagate at a plane interface between a solid and a fluid (surface waves, transverse lateral wave in the fluid, lateral wave in the solid, etc.). For instance, as shown in Fig. 2 inside the solid, a lateral wave front connects the longitudinal and the transverse refracted wave fronts. Note that the orientation of this lateral wave front depends on the ratio of shear velocity to longitudinal velocity, which is similar here to the ratio of velocity in water to longitudinal velocity.

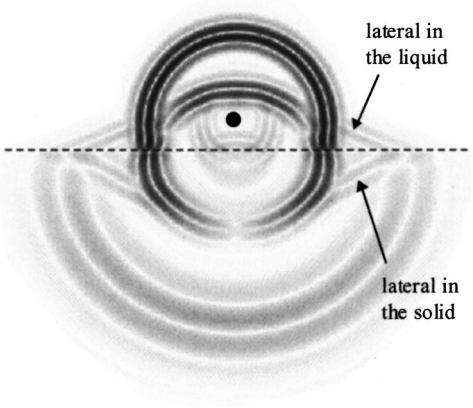


FIG. 2. Snapshot of the wave propagation, in the fluid/solid case, after the longitudinal refracted wave front has disconnected from the incident wave front.

2. Time-of-flight of the lateral wave

Let us consider a source and a receiver both at a distance d from the interface (Fig. 3) between the two media, separated by a distance r. The lateral wave exists at the receiver only if $r > 2d\tan(\theta_c)$, so that path BC exists.[6] Henceforth, we suppose that r always verifies this condition. When does the lateral wave arrive first at the receiver? For the lateral wave, the theory predicts that the TOF corresponds to the geometrical path *ABCD* (Fig. 3), traveled with velocity c_w on part *AB* and *CD* and with velocity c_l on part *BC*. It corresponds to the path cited by several authors[3–5] using the axial transmission technique. This TOF equals[6]

$$t_{\text{lateral}} = \frac{2d\cos(\theta_c)}{c_w} + \frac{r}{c_l}. \tag{5}$$

The TOF of the direct and the reflected waves are, respectively,[6]

$$t_{\text{direct}} = \frac{r}{c_w}, \tag{6}$$

$$t_{\text{reflected}} = \frac{\sqrt{r^2 + (2d)^2}}{c_w}. \tag{7}$$

It can be seen from these expressions that r must be large enough, so that the lateral wave, propagating along the surface at $c_l > c_w$ arrives first. The corresponding condition $t_{\text{direct}} - t_{\text{lateral}} > 0$ therefore yields a value $r_{\min}$, so that for r

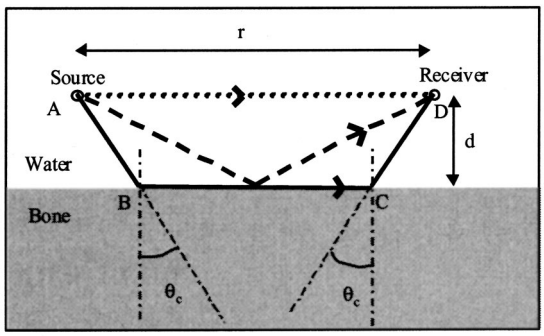


FIG. 3. Ray paths for the direct (···), reflected (---), and lateral wave (—).

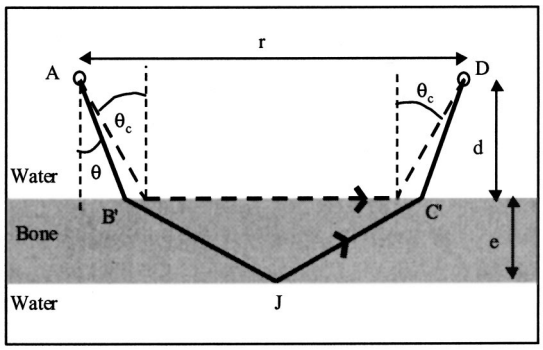


FIG. 4. Ray paths for the lateral wave (---) and the wave reflected from the bottom of the plate (—).

$> r_{\min}$, the TOF of the lateral wave is less than that of the direct wave.

B. Extension to the case of an immersed bone plate

In the case of an infinite solid plate immersed in water, we apply the ray theory to the different waves present in the case of the fluid/solid interface. We suppose that the lateral wave still exists at the first water/solid interface, excited by incident rays with an angle of incidence θ greater than θ_c, and has properties identical to those in the case of a single interface between semi-infinite media (ray path, time-of-flight). For incident θ smaller than θ_c, incident rays penetrating the plate are reflected on the solid/fluid interfaces and give rise to longitudinal reflected rays and transverse reflected rays, which are then refracted on the solid/fluid interface, and so on. Those reflections and refractions obey Snell–Descartes laws, with velocity c_t along the path of the transverse waves, and velocity c_l along the path of the longitudinal waves. Several paths, involving multiple reflections inside the plate, go from the source to a given receiver. In addition to the path *ABCD* of the lateral wave, we will consider here the path $AB'JC'D$ (Fig. 4), which involves only longitudinal waves, and corresponds to a single reflection on the bottom of the plate. This path is the fastest path inside the plate from the source to the receiver. For a given receiver at distance r from the source, the angle of incidence θ of a ray that follows path $AB'JC'D$ and the associated TOF $t_{AB'JC'D}$ verify the following equations, which can be solved numerically:

$$t_{AB'JC'D} = \frac{2d}{c_w} \times \frac{1}{\cos(\theta_c)} + \frac{2e}{c_l} \times \frac{1}{\sqrt{1 - \left(\dfrac{\sin(\theta)}{\sin(\theta_c)}\right)^2}}, \tag{8}$$

$$r = 2d\tan(\theta) + 2e\,\frac{\sin(\theta)}{\sin(\theta_c)} \times \frac{1}{\sqrt{1 - \left(\dfrac{\sin(\theta)}{\sin(\theta_c)}\right)^2}}. \tag{9}$$

The difference in TOF $t_{AB'JC'D} - t_{\text{lateral}}$ depends on the distance r and the thickness e of the plate. A numerical resolution of Eq. (8) shows that $t_{AB'JC'D} > t_{\text{lateral}}$ for any values of e and r, which implies that the TOF is always shorter for

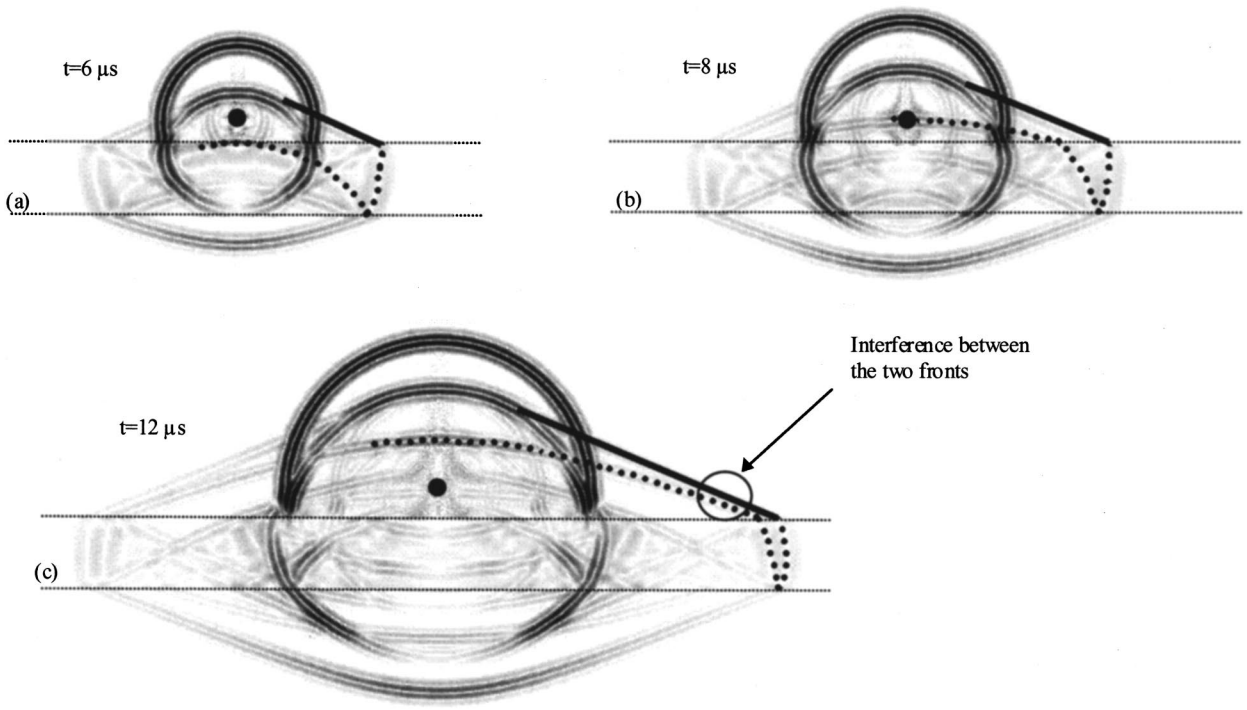


FIG. 5. Snapshots of the wave propagation in the case of a bone plate, 6 mm thick, at three different times [(a) $t=6\ \mu$s, (b) $t=8\ \mu$s, and (c) $t=12\ \mu$s]. Interference between the lateral wave (—) and the wave reflected from the bottom of the plate (---) is seen on snapshot (c).

the lateral wave. However, for small plate thickness e or large distance r, the time delay between the lateral wave and the wave reflected from the bottom is small, as can be inferred from ray paths of Fig. 4. When this time delay becomes shorter than the lateral wave pulse duration, there is interference between the two waves, as seen on snapshots of the wave propagation at three different times (Fig. 5). In Fig. 5(c), the difference in TOF between the two highlighted wave fronts (lateral wave front and wave front reflected from the bottom) is of the order of a period T, and the two fronts interfere with each other.

We shall demonstrate in the following that this geometrical interpretation does not hold in the case of very thin plates for which plate modes have to be considered.

C. Measurement of the TOF of the lateral wave

As indicated above, the TOF verify the following equations:

$$t_{\text{direct}}-t_{\text{lateral}}>0 \quad (\text{for } r>r_{\min}), \tag{10}$$

$$t_{AB'JC'D}-t_{\text{lateral}}>0 \quad (\text{for any } r). \tag{11}$$

Those indicate that the lateral wave has the shortest TOF. Yet, depending on the time criterion used to detect the FAS (for example, extremum, zero crossing or cross correlation, discussed further in Sec. III B), the conditions to actually obtain an unbiased value of the TOF of the lateral wave have to be detailed. Indeed, the differences in TOF ($t_{\text{direct}}-t_{\text{lateral}}$ and $t_{AB'JC'D}-t_{\text{lateral}}$) have to be large enough, as discussed below, so that the detected part of the lateral wave does not interfere with the following waves. In case of such interference, the estimate of the TOF will be erroneous, with an error depending on the time criterion. Typically, if the detected part is the first extremum or the first zero crossing of the lateral wave, which is detected within a period after the beginning of the signal, the conditions to actually measure an unbiased TOF are

$$t_{\text{direct}}-t_{\text{lateral}}>\varepsilon, \tag{12}$$

$$t_{AB'JC'D}-t_{\text{lateral}}>\varepsilon, \tag{13}$$

where ε is of the order of the mean time period of the signal (typically 1 μs for our simulations at 1 MHz). The resolution of Eq. (12) and Eq. (13) leads us to define a zone in the (r,e) plane in which the measured TOF actually corresponds to the TOF of the lateral wave. Equation (12) yields a value $r_{\min}(\varepsilon)$, so that for $r>r_{\min}(\varepsilon)$, the direct wave does not interfere with the detected part of the lateral wave. Table I reports the distance $r_{\min}(\varepsilon)$, for different values of ε. Figure 6 plots $t_{AB'JC'D}-t_{\text{lateral}}$ as a function of distance r, for several plate thickness' e: when thickness e decreases for a given distance r, or when distance r increases for a given thickness e, the difference $t_{AB'JC'D}-t_{\text{lateral}}$ vanishes toward zero. Thus Eq. (13) involves that for a fixed thickness e, distance r must be inferior to a certain distance $r_{\max}(\varepsilon,e)$, and for a fixed distance r, thickness e must be superior to a certain value $e_{\min}(\varepsilon,r)$. Therefore, these two conditions allow us to predict the existence of a limited range of values r and e so that the measured TOF corresponds exactly to the TOF of the lateral

TABLE I. Distance $r_{\min}$ for different values of ε.

ε (μs)	0	0.5	1	1.5
$r_{\min}$ (mm)	5.9	7.1	8.3	9.5

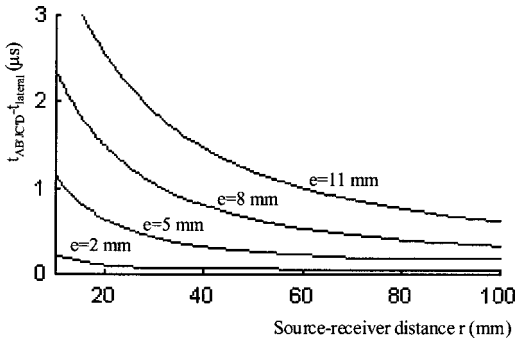


FIG. 6. Difference in TOF $t_{AB'JC'D}$-t_{lateral} as a function of source–receiver distance r, for several plate thicknesses e.

wave, which does not interfere with the following waves. The exact limits of these ranges depend on the detection criterion. Solutions of Eq. (12) and Eq. (13) are presented in the results part.

III. SIMULATIONS METHODS

A. Software

We used a commercially available 2D simulation software (Wave2000 Pro, CyberLogic, Inc., New York, USA) which computes a numerical solution to the 2D elastic wave propagation, based on a finite difference method.

Osteons, the largest structures of cortical bone, have typical dimensions of about 250 μm.[10] At 1 MHz, the wavelength of ultrasound is about 4 mm in cortical bone, much greater than the osteons. Therefore, we modeled cortical bone as a homogeneous solid. Furthermore, we neglected anisotropy and absorption in bone, and thus characterized the cortical bone by its density (1850 kg/m^3), its longitudinal bulk velocity (c_l=4000 m/s), and its transverse bulk velocity (c_t=1800 m/s), consistent with typical values found in the literature.[11] The bone specimen is immersed in nondissipative water, with a density of 1000 kg/m^3 and bulk velocity c_w=1490 m/s.

The three typical wavelengths, at 1 MHz, are λ_l =4 mm (longitudinal bulk wavelength in bone), λ_t =1.8 mm (transverse bulk wavelength in bone), and λ_w =1.5 mm (bulk wavelength in water). The space increment was $\Delta x=\Delta z=0.025$ mm (corresponding to 60 grid points per shortest wavelength) in the case of the single water/bone interface, and $\Delta x=\Delta z=0.05$ mm (corresponding to 30 grid points per shortest wavelength) for the case of immersed

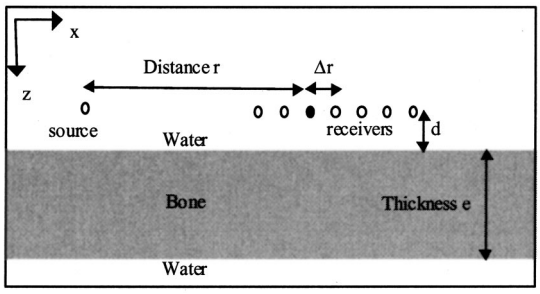


FIG. 7. Typical geometrical configuration for a simulation of the axial transmission technique.

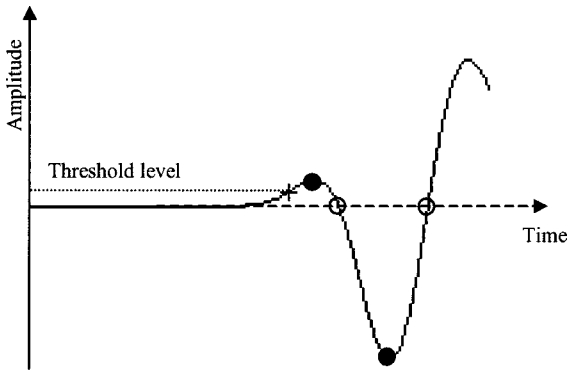


FIG. 8. Illustration of different time detection criteria (+, threshold-time; ○, zero-crossings; ●, extrema).

bone plates. The time increment was automatically deduced by the software from the von Neumann stability condition,[12] which led, respectively, to sampling frequency of about 180 MHz and 90 MHz.

We used a spherical source that transmits a 1 MHz broadband pulse, and several pointlike receivers. The typical set of geometrical parameters characterizing each simulation is shown in Fig. 7. The receivers measure displacement amplitudes in the z direction. The source and receivers were at a distance d=2 mm from the interface. The distance r ranged from about 4 mm to 100 mm, and thickness e ranged from 0.5 mm to 12 mm (values e=0.5,1,2,3,...,11,12 mm) for the plates. The interdistance between two adjacent receivers was Δr=2 mm in the case of a single interface between two semi-infinite media, and was Δr=3 mm for bone plates.

B. Signal processing

For all the simulated signals, the amplitudes are in linear arbitrary units, and are directly comparable from one simulation to another. The aim of our simulations was to measure the velocity of the first arriving signal, the so-called apparent SOS. This was achieved by deriving the time delay between the signals arriving at two adjacent receivers, either by measuring and subtracting times of flight (TOF), or by calculating cross correlations. The cross correlations involved windowed signals, with a 3 μs duration rectangular window centered for each signal on the first arriving pulse. Cross correlation were only calculated when the FAS consisted of pulses temporally well separated from the subsequent signal. Both these methods yield a time interval Δt. Several time

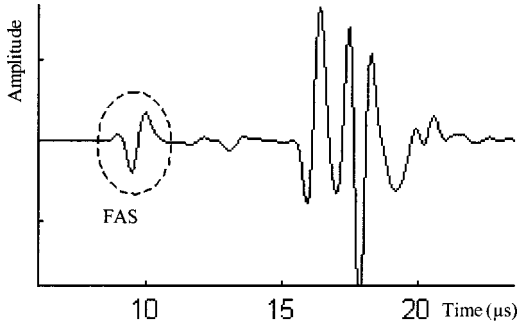


FIG. 9. Typical radio-frequency signal, with the FAS temporally separated from the subsequent signal.

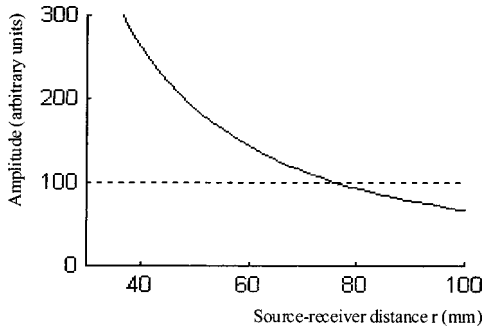


FIG. 10. Amplitude of the first extremum of the lateral wave, as a function of source–receiver distance r.

criteria, which are illustrated in Fig. 8, were used to define the TOF at one receiver, including extrema, zero crossings or threshold-based time criteria. Extrema were determined using a parabolic interpolation between three samples around the extrema. The threshold-based TOF was defined as the time for which the absolute value of the radio-frequency signal reaches a predetermined threshold level for the first time, and was estimated using a linear interpolation between the two samples around the threshold level. The zero-crossing TOF were determined similarly using a linear interpolation between the two samples around the zero crossings. For each measurement of Δt, the apparent SOS was derived for a pair of receivers, separated by a distance Δr, with the following equation:

$$\text{apparent SOS} = \frac{\Delta r}{\Delta t}. \tag{14}$$

IV. RESULTS

A. Interface between two semi-infinite media

Figure 9 shows a typical RF signal, with the lateral wave arriving first, temporally well separated from the subsequent signal. The amplitude of the first extremum of the lateral wave, as a function of distance r, is reported in Fig. 10. Figure 11 shows the apparent SOS plotted against distance r between the source and the pair of receivers for several detection criteria.

First, we notice that threshold-based apparent SOS values are consistently lower (of the order of a few %) than c_l, and depends strongly on the threshold level. Apparent SOS values derived with a threshold level of 100 (arbitrary units) are only calculated for distances $r<75$ mm, since the magnitude of the first extremum decreases as r increases, and becomes smaller than the threshold level of 100 for distances larger than 75 mm.

Second, for measurement based on time criteria independent of the amplitude of the signal (extremum, zero crossing, cross correlation), the apparent SOS value was found to be c_l, with an accuracy better than 0.1%, for distances r typically larger than about $10.\lambda_l$ (40 mm). For shorter distances, the apparent SOS is consistently lower than c_l, and depends on the time criterion.

B. Bone plates

We intend here to define different zones in the (r,e) plane, derived from the measured TOF of the FAS. We define the *zone of observability of the lateral wave* as the zone for which the measured TOF for a given distance r is independent of plate thickness and equals that found on the semi-infinite bone medium. Figures 12(a) and (b) show different zones in the (r,e) plane, derived from the simulated TOF, for two detection criteria. The $(r-e)$ plane has been divided into three zones (white, gray, and black). The white zone corresponds to the zone of observability defined above. In the black zone, the measured TOF is independent of plate thickness, and corresponds to the TOF of the direct wave that arrives first. In the gray zone, the measured TOF is a function of both e and r. The superimposed curve is obtained by solving numerically the following equation:

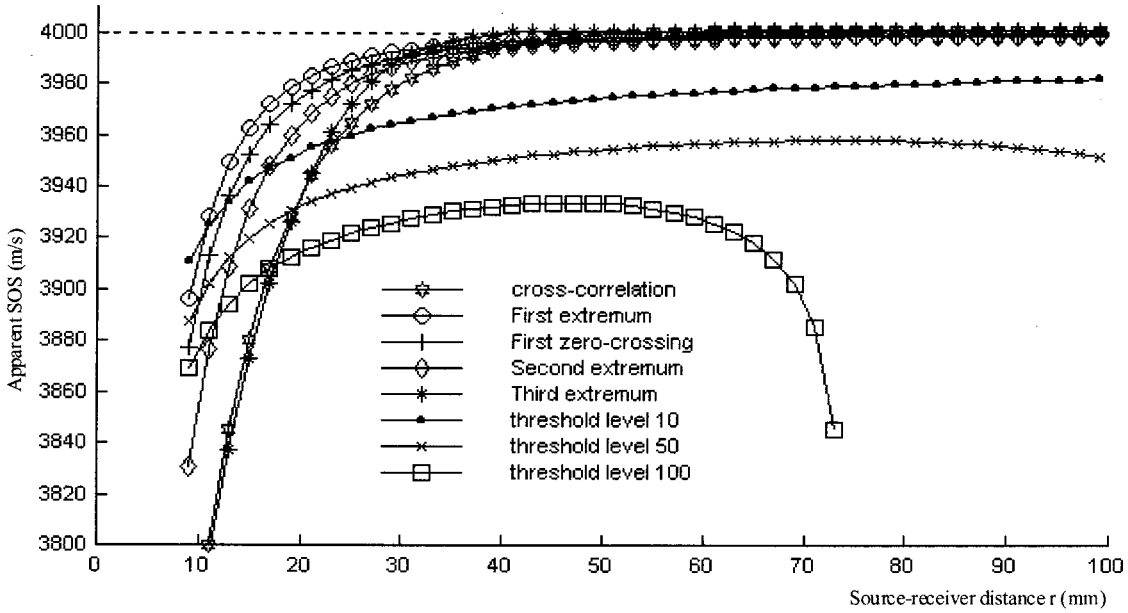


FIG. 11. Apparent SOS as a function of source–receiver distance, for several time detection criteria.

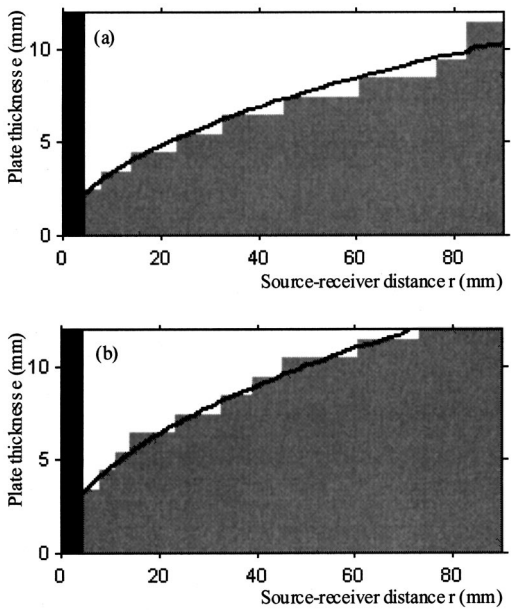


FIG. 12. (r,e) planes divided in white, gray, and black zones, for two differents time detection criteria, and superimposed solution of Eq. (15). (a) Detection of the first extremum, $\varepsilon = 0.6\ \mu$s. (b) Detection of the second extremum, $\varepsilon = 1\ \mu$s.

$$t_{AB'JC'D} - t_{\text{lateral}} = \varepsilon \tag{15}$$

with ε chosen so that the calculated curve fits the separation between the gray zone and the white zone. This illustrates the good agreement between theoretical and simulated results. ε values are, respectively, $\varepsilon = 0.6\ \mu$s and $\varepsilon = 1.0\ \mu$s for the detection of the first extremum and the second extremum.

Figure 13 shows rf signals for different plate thickness', recorded at a distance $r = 20$ mm and $r = 30$ mm. Those values of distance r are typical of those which were used *in vivo* at sites such as the radius or tibia.[13,14] The thickness values presented in Fig. 13 include values of cortical thickness of skeletal sites such as radius, tibia, femur, metacarpal, phalanx.[7,15,16] The interference between different waves, clearly depicted in Figs. 13(b) and (e), results in a much lower signal amplitude for $e = 3$ mm than $e = 0.5$ mm or $e = 5$ mm. In Fig. 14, the apparent SOS is plotted as a function of plate thickness, for several distances r and several time detection criteria. For plates thick enough (typically $e > \lambda_l$), the apparent SOS is independent of plate thickness, in agreement with Fig. 12, and is exactly identical to that measured in the case of the single interface between semi-infinite media. For thinner plates (typically $e \leqslant \lambda_l$), the pattern of variation of the apparent SOS depends on both the detection criterion and the distance r. In particular, whereas the apparent SOS derived from a threshold time consistently decreases with thickness, the apparent SOS derived from the other criteria present a local maximum for a thickness of the order of λ_l. For very thin plates (typically $e \leqslant \lambda_l/8$) and for time detection criteria independent of the amplitude (such as extrema or zero crossings), the apparent SOS approaches the velocity of the S_0 Lamb mode ($c_{S0} = 3215$ m/s). The shape of the curves derived from threshold time criteria are very similar to those published by Njeh *et al.*[7] for *in vitro* measurements.

Time delay Δt was not calculated by cross correlation for bone plates, as the FAS did not consist in a single pulse, separated from the subsequent signal.

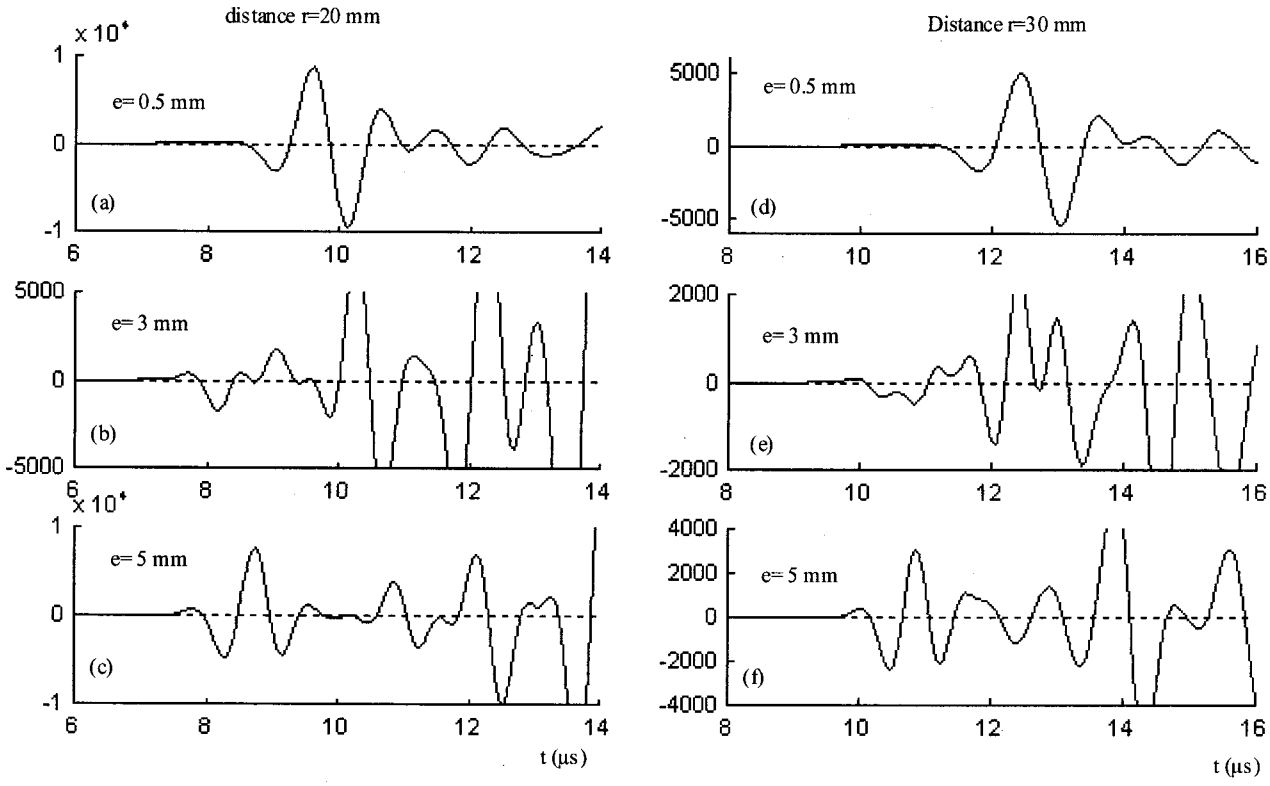


FIG. 13. rf signals for three plate thickness $e = 5$ mm, $e = 3$ mm, and $e = 0.5$ mm, recorded at distances $r = 20$ mm and $r = 30$ mm.

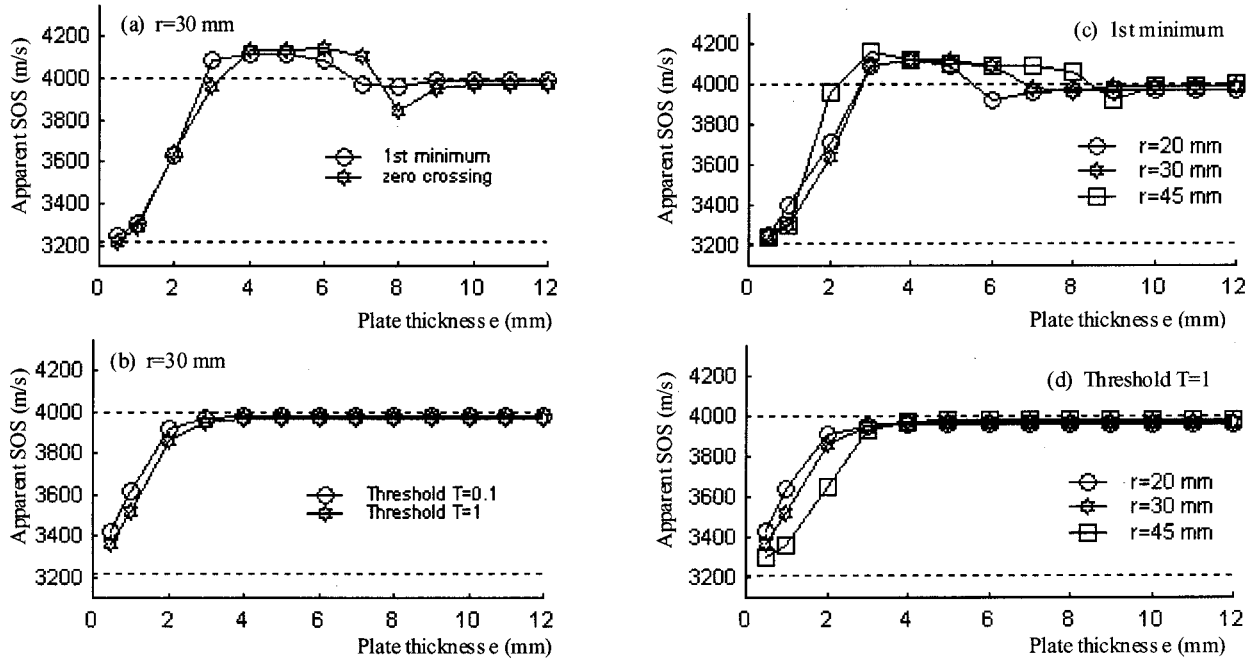


FIG. 14. Apparent SOS as a function of plate thickness, for several detection criteria and distances r.

V. DISCUSSION

The case of an interface between two semi-infinite media, developed theoretically and experimentally by Camus *et al.*,[6] have been simulated in our work to study the accuracy of the measurement as a function of the detection algorithm, and to be used as a reference to compare to results obtained with bone plates, as a function of plate thickness.

A. Interface between two semi-infinite media

For a single plane interface between water and bone, the first arriving signal (FAS) measured with the so-called axial transmission technique corresponds to the lateral wave. The apparent SOS, defined as the velocity of the FAS, equals the longitudinal bulk velocity of the bone with an accuracy of 0.1% for source–receiver distances larger than about $10\lambda_l$, and for time detection independent of the amplitude of the signal. The apparent SOS values derived from threshold-based criteria are dependent on the amplitude of the signal, in contrast to apparent SOS values derived from other time criteria including extrema, zero crossings or cross correlation. The underestimated values of apparent SOS yielded by threshold-based detection correspond to a decrease in amplitude of the lateral wave with the distance r.[6,9] The part of the lateral wave that reaches the threshold differs between two receivers, as the amplitude has decreased from one receiver to the other. This leads to a value Δt larger than that due to propagation delay only. As this bias depends on several parameters, such as the amplitude of the signal relatively to the threshold and the way the wave is attenuated from one receiver to the other, a threshold-based detection is inadequate to accurately and unambiguously measure the SOS of the lateral wave. Moreover, attenuation due to material absorption, neglected in this work, would increase this bias for *in vivo* measurements. But why is the apparent SOS derived from amplitude-independent criteria different from c_l for small values of r? As the apparent SOS measured for large distances r and by different signal processing techniques (TOF detection and cross correlation) all tend towards c_l, as predicted by the theory, with an accuracy better than 0.1%, the answer to this question probably has a physical origin, rather than being caused by numerical artifacts from the simulations or signal processing errors. For short distances (typically $r<10\lambda_l$), it has to be noted that the difference between the apparent SOS and c_l is only of the order of a few percent. The discrepancy between the apparent SOS and c_l may come from the fact that the analytical expression of the lateral wave, which states that the wave propagates along the surface with velocity c_l, is given by an asymptotic theory that is valid only for large distance r.[9,17] Furthermore, Donato[17] has pointed out that the validity of this asymptotic theory is different whether the high velocity medium is a fluid or a solid: whereas for a fluid high velocity medium, the distance r has to be of the order of one or two wavelengths for the asymptotic expression to be valid, it has to be several wavelengths in the case of a solid high velocity medium. We checked this statement with additional simulations achieved for a fluid high velocity medium of velocity $c_l=4000$ m/s and density 1850 kg/m^3, as shown in Fig. 15. The apparent SOS reaches c_l with an accuracy of 0.1% for a distance r of about $4\lambda_l$, in contrast to about $10\lambda_l$ for a solid high velocity medium. The fact that the apparent SOS values derived from amplitude-independent method all converge towards the same value c_l, but differ slightly from each other for small distances r, suggests that the shape of the lateral wave changes before it eventually stabilizes for long distances r. This change in the shape of the lateral wave with propagation seems to be different for the fluid/fluid case compared to the fluid/solid case. Indeed, in the fluid/fluid case, the apparent SOS is consistently larger than c_l, whereas it is consistently lower in the fluid/solid case. Addi-

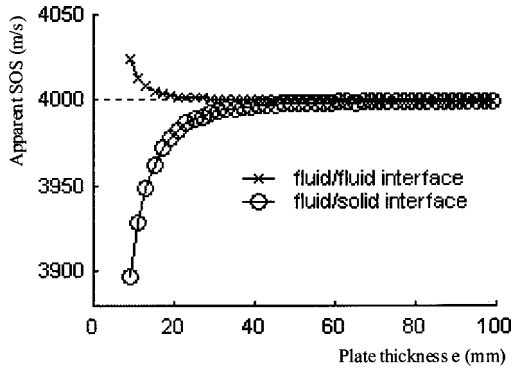


FIG. 15. Apparent SOS as a function of source–receiver distance, for the cases of fluid and solid high velocity media.

tional simulations on test media such as aluminum ($c_l = 6400$ m/s, $c_t = 3040$ m/s, $\rho = 2700$ kg/m^3) and Plexiglas ($c_l = 2700$ m/s, $c_t = 1100$ m/s, $\rho = 1150$ kg/m) lead exactly to the same conclusions (not shown here).

B. Bone plates

For plates thick enough (thickness typically $e > \lambda_l$), and for a limited range of values of r depending on the thickness, the FAS still corresponds to the lateral wave, and the apparent SOS is exactly the same as that measured in the case of a single interface between two semi-infinite media. This validates the use of the axial transmission technique to measure the longitudinal bulk velocity in plates, provided that the plate thickness is greater than a certain thickness e_{limit} and that the source–receiver distance is not too large. The thickness e_{limit} depends on the source–receiver distance r, but is at least greater than the longitudinal wavelength in the plate. This result is correctly explained in terms of time-of-flight predicted by the ray theory. Practically, the choice of a time detection criterion yields a value of time delay ε, which can then be used in Eq. (15) to predict the range of plate thickness e and range of distance r that allow a correct measure of the TOF of the lateral wave. These ranges yield the zone of observability of the lateral wave, as shown in Fig. 12. In the zone of observability, the errors in apparent SOS that may result from inappropriate detection algorithm, or too small distance r, are identical to the case of the single interface, and will not be discussed here.

For thin plates of thickness $e < \lambda_l/4$, the ray theory is no longer a valid description of the phenomena, and a modal approach of the signal is necessary. The FAS is nearly nondispersive, as its shape does not change with distance r [Figs. 13(c) and (f)], and has an apparent SOS around 3200–3250 m/s (Fig. 14). This FAS is therefore likely to correspond to the lowest order symmetrical mode (S_0 Lamb mode), which is nondispersive for very thin plate ($e \ll \lambda_l$) and has an asymptotic phase velocity, equal to its group velocity, given by[18]

$$c_p = 2c_t \sqrt{1 - \left(\frac{c_t}{c_l}\right)^2} = 3215 \text{ m/s}. \tag{16}$$

In a frequency/wave-number diagram (such as Figs. 16 and 17), a mode of propagation is characterized by a dispersion

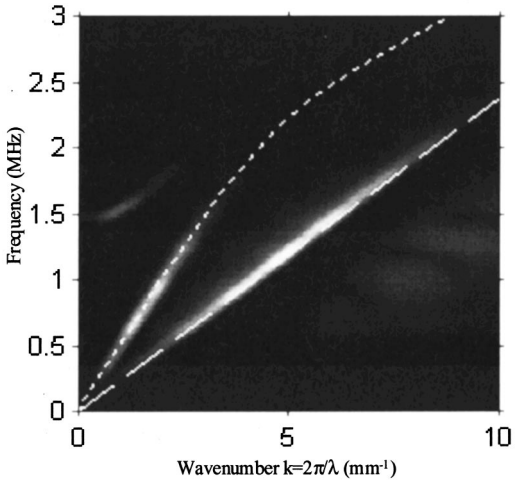


FIG. 16. Theoretical dispersion curves for a free bone plate of thickness $e = 0.5$ mm, superimposed over the frequency/wave-number diagram obtained from simulated (r,t) diagrams ($\cdots$, S_0 mode; ---, direct and reflected wave).

curve, that gives the relationship between the temporal frequency of the wave and its wavelength (wave-number $k = 2\pi/\text{wavelength}$). The theoretical dispersion curves for a free bone plate of thickness $e = 0.5$ mm are shown in Fig. 16, superimposed over the frequency/wave-number diagram obtained by performing a 2D fast-Fourier transform (FFT) of the distance–time diagram. This figure shows that the whole signal is due to the S_0 mode, the direct wave and the reflected wave. As the S_0 mode propagates faster than the direct and reflected wave, it necessarily corresponds to the FAS, provided that it has outdistanced the direct wave, which is obtained for a distance r typically larger than 10 mm.

For plates of intermediate thickness, such as $\lambda_l/4 < e < \lambda_l$, the apparent SOS depends on thickness e, but also on the detection criterion and on distance r. The shape of the FAS dramatically changes for intermediate thicknesses, compared to thick and thin plates, as seen in Fig. 13. Moreover this shape strongly depends on the propagation distance r, as

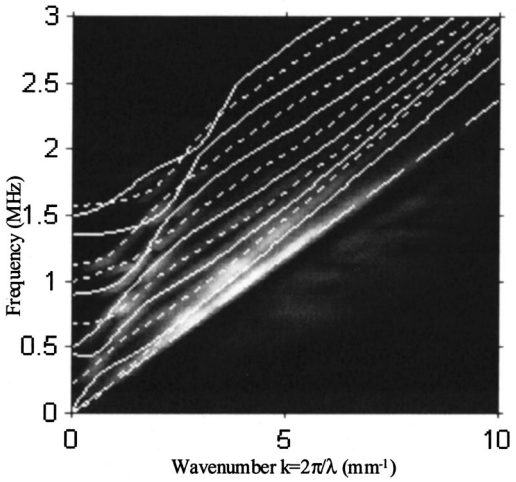


FIG. 17. Theoretical dispersion curves for a free bone plate of thickness $e = 4$ mm, superimposed over the frequency/wave-number diagram obtained from simulated (r,t) diagrams ($\cdots$, first six symmetrical modes; —, first six antisymmetrical modes; ---, direct wave).

seen in Figs. 13(b) and (e). Figure 17 shows that several modes contribute to the whole signal for a thickness $e = \lambda_l$. The FAS probably results from a complex pattern of interference between different waves, which is not fully understood yet. Destructive or constructive interference may occur, resulting in a complex wave form [Figs. 13(b) and (e)], with an amplitude of FAS that can be one order of magnitude lower than for thick and thin plates. For those intermediate thickness, what does the apparent SOS correspond to? For thickness e of the order of λ_l, the FAS may be interpreted as an interference between the lateral wave and the wave reflected from the bottom of the plate, whereas for thickness of the order of $\lambda_l/4$, it may be interpreted as an interference between plate modes. The nature of the FAS continuously changes between those two limits. Eventually, in this range of intermediate thickness, the FAS always corresponds to interference between different waves propagating at different velocities, and the measurement of a TOF yields an apparent SOS which is neither a phase nor a group velocity, and has no other physical meaning than a signal velocity. In particular, the apparent SOS corresponds neither to the lateral wave velocity nor to a single pure mode phase or group velocity. For this reason, the term apparent SOS was preferred to the term SOS, often used by others in axial transmission experiments.[3,4,19] Practically, for intermediate thickness, as the amplitude and interference pattern of the FAS strongly depends on thickness and distance r, it is difficult to define an automated time criterion that yields an unambiguous measurement of the TOF. For instance, at distance $r = 30$ mm, the second zero crossing does not have the same meaning for thickness $e = 5$ mm, $e = 3$ mm, and $e = 0.5$ mm (Fig. 13). It has also to be noted that in our simulations the detection of any extremum was always unambiguous as the signal was free of noise, which allowed us to detect extrema level as low as 0.1 or 1 units, compared to typical amplitudes of a few hundreds for the lateral wave. For real experiments, whatever the detection algorithm, the detection of the FAS requires the definition of a threshold level, conditioned by the level of random noise. Depending on thickness and distance r, the amplitude of the first arriving wave may be lower than the level of noise, and the interpretation of the detected signal may therefore become more difficult.

For practical experiments, the choice of a particular detection algorithm may depend on the kind of information that one looks for. If accurate measurements are desired, the chosen time detection criteria have to be independent of the amplitude (extrema, zero crossing), unlike TOF derived from threshold-based time detection, which can yield bias of the order of a few percent. If the apparent SOS serves as an indirect way to evaluate cortical thickness, it may be more relevant to use measurements derived from threshold-based time detection, even biased, as they lead to a constant relationship between apparent SOS and thickness, that is the apparent SOS consistently decreases with cortical thickness. The dependence of the apparent SOS on plate thickness obtained from our simulations, with a threshold-based time detection, is similar to that published from *in vitro*[7] and *in vivo*[15] experiments at 1.25 MHz. From our results, this relationship may not only be due to the variation in thickness of the measured material, but also to experimental conditions (source–receiver distance, detection criteria). Therefore, one has to be cautious when comparing values obtained from different devices and interpreting clinical data. Particularly, the detection methods for commercially available devices using axial transmission technique are not made explicit in published works.[15,20,21] As two different devices may give two different measurements on a given bone, because of possible different time criteria or different source–receiver distances, it could therefore be rather awkward to compare clinical measurements obtained from different devices.

As bone is heterogeneous and bone properties may vary with depth, this study raises the question of the depth of bone which is probed with the axial transmission technique. This is particularly important when bone resorption mostly occurs at the inner part of the cortical shell, as in osteoporosis and ageing.[8] The technique must then be able to provide quantitative measurements reflecting not only bone changes of the very superficial layers, but also changes in deeper layers. Toward this goal, we are currently simulating the effect of a velocity gradient at the bone surface in the case of a single interface, as well as the effect of porosity throughout the cortical shell. The present study performed on homogeneous bone plates demonstrates the existence of a critical bone thickness e_{limit} such that for $e < e_{\text{limit}}$, the apparent SOS differs significantly from the longitudinal bulk velocity. This thickness e_{limit} is not only dependent on bone properties, but also on the measurement technique (source–receiver distance, detection criterion). The lower values of e_{limit} are of the order of the longitudinal wavelength in bone. A complete answer to that question in the case of a heterogeneous medium of finite thickness is not clear yet and further work is necessary to determine a penetration depth of the axial transmission technique in bone.

C. Applicability of the model

Many hypotheses had to be made to simulate axial transmission technique in our work. The source in our simulations was modeled by a spherical source, whereas actual transducers have finite size and do not produce spherical wave. Yet, we chose to use a spherical source as this kind of ideal source produces all possible angles of incidence, and therefore excites all the possible waves that may propagate in axial transmission. Indeed, the angle of incidence θ needed to excite a wave propagating along the surface of bone with a velocity v_{bone} is given by $\sin(\theta) = v_{\text{fluid}}/v_{\text{bone}}$, for a source located in a fluid of velocity v_{fluid}. From our results, the velocity of the FAS ranged between approximately 3000 ms^{-1} and 4000 ms^{-1}, which corresponds to incidence angle ranging between approximately 22° and 30°. Practically, our results are therefore applicable to real transducers that at least transmit some acoustic energy with incidence angle between approximately 22° and 30°, which is obtained on current commercially available systems by tilting the transducers relatively to the surface of bone. Further hypothesis have been made on the nature of bone, which may have an impact on axial transmission measurement. First, attenuation in bone and fluid (modeling gel and soft tissues) has been neglected.

This may be justified in the fluid as the acoustic paths are the same for waves arriving at two different receptors. In bone, attenuation may result in a change of the pulse shape because of possible frequency-dependent attenuation. However, cortical bone has been shown to be nearly nondispersive in relatively large frequency ranges,[10] in contrast with the situation for cancellous bone.[22] Therefore the change of the pulse shape has been considered negligible. Nevertheless, the decrease in amplitude of the FAS due to attenuation (frequency independent) may be a parameter of interest to characterize bone status, which is currently investigated in our laboratory. Second, anisotropy of bone has been neglected. As human cortical bone has been shown to be transversely isotropic,[10] the velocity measured in axial transmission (for cortical shell thick enough) should correspond to the velocity along the axis of bone. For thinner cortical shells, the nature of propagated waves is probably different from Lamb modes obtained for isotropic plates, but this does not modify our conclusion concerning a change in the nature of the wave when the cortical thickness decreases. We are currently investigating further 2D simulations, taking into account the difference between axial and radial velocity in cortical shells.

The main limitation of our work concerns the modeling of a curved cortical shell by a plate and the applicability of our simulations results to real simulations. However, as shown experimentally, the dependence on thickness of the apparent SOS by axial transmission is similar for plates and tubes,[7] the apparent SOS decreasing with thickness in both cases. Moreover, measurements on tubular thick shell lead to the longitudinal bulk velocity of the material, as it does for a thick plate of the same material.[7] *In vivo* measurements also lead to velocity values consistent with longitudinal bulk velocity in cortical bone.[3] These results justify our plate model as a first approach to the problem. Nevertheless, three-dimensional simulations, taking into account a more realistic geometry for bone, are required to better understand the wave propagation that occurs in *in vivo* measurements.

VI. CONCLUSION

We have shown that the axial transmission technique allows measurements of the longitudinal wave velocity in the case of bone plates, provided that the plate thickness is at least greater than the longitudinal wavelength in bone λ_l, and that the source–receiver distance is not too large. For thin plates, of thickness less than about $\lambda_l/4$, the technique yields the velocity of the nondispersive lowest order symmetrical plate mode (S_0 Lamb mode). For intermediate thickness, the apparent SOS is dependent on the method of measurement, and is therefore not intrinsic to the measured material.

ACKNOWLEDGMENTS

We gratefully acknowledge J. J. Kaufman for helpful discussions about Wave2000 Pro.

[1] D. Hans, A. M. Schott, J. L. Sebert, C. Cormier, P. Kotzki, P. D. Delmas, J. M. Pouilles, G. Breart, and P. J. Meunier, ''Ultrasonographic heel measurements to predict hip fracture in elderly women: The epidos prospective study,'' Lancet **348**, 511–514 (1996).
[2] I. Siegel, G. T. Anast, and T. Melds, ''The determination of fracture healing by measurement of sound velocity across the fracture site,'' Surg. Gynecol. Obstet. **1958**, 327–332 (1958).
[3] R. Barkmann, E. Kantorovich, C. Singal, D. Hans, H. K. Genant, M. Heller, and C. C. Glüer, ''A new method for quantitative ultrasound measurements at multiple skeletal sites,'' J. Clinical Densitometry **3**, 1–7 (2000).
[4] D. Hans, S. Srivastav, C. Singal, R. Barkmann, C. Njeh, E. Kantorovich, C. C. Glüer, and H. K. Genant, ''Does combining the results from multiple bone sites measured by a new quantitative ultrasound device improve discrimination of hip fracture?,'' J. Bone Miner. Res. **14**, 644–651 (1999).
[5] G. Lowet and G. Van der Perre, ''Ultrasound velocity measurement in long bones: Measurement method and simulation of ultrasound wave propagation,'' J. Biomech. **29**, 1255–1262 (1996).
[6] E. Camus, M. Talmant, G. Berger, and P. Laugier, ''Analysis of the axial transmission technique for the assessment of skeletal status,'' J. Acoust. Soc. Am. **108**, 3058–3065 (2000).
[7] C. Njeh, D. Hans, C. Wu, E. Kantorovich, M. Sister, T. Fuerst, and H. K. Genant, ''An *in vitro* investigation of the dependence on sample thickness of the speed of sound along the specimen,'' J. Med. Eng. Phys. **21**, 651–659 (1999).
[8] C. Njeh, X. G. Cheng, J. M. Elliot, and P. J. Meunier, ''Bone, bone disease and bone quality,'' *Quantitative Ultrasound. Assessment of Osteoporosis and Bone Status* (Martin Dunitz, London, 1999), pp. 1–20.
[9] L. Brekhovskikh and O. Godin, *Acoustics of Layered Media II: Point Sources and Bounded Beams. Wave Phenomena* (Springer Verlag, Berlin, 1992).
[10] P. Nicholson and M. L. Bouxsein, ''Ultrasonics studies of cortical bone *in vitro*,'' *Quantitative Ultrasound. Assessment of Osteoporosis and Bone Status* (Martin Dunitz, London, 1999), pp. 177–193.
[11] S. S. Mehta and P. P. Antich, ''Measurement of shear-wave velocity by ultrasound critical-angle reflectometry (ucr),'' Ultrasound Med. Biol. **23**, 1123–1126 (1997).
[12] R. A. Stephen, ''A review of finite difference methods for seismo-acoustics problems at the seafloor,'' Rev. Geophys. **26**, 445–458 (1988).
[13] G. Lowet, G. Van der Perre, and J. Lammens, ''Monitoring of bone consolidation by ultrasound velocity measurement,'' IEEE Eng. Med. Biol. Soc. **5/7** (14th Annual International Conference) (1992).
[14] J. M. Orgee, H. Foster, E. V. McCloskey, S. Khan, G. Coombes, and J. A. Kanis, ''A precise method for the assessment of tibial ultrasound velocity,'' Osteoporosis Int. **6**, 1–7 (1996).
[15] H. Sievänen, S. Cheng, S. Ollikainen, and K. Uusi-Rasi, ''Ultrasound velocity and cortical bone characteristics *in vivo*,'' Osteoporosis Int. **12**, 399–405 (2001).
[16] S. Prevrhal, T. Fuerst, B. Fan, C. Njeh, D. Hans, M. Uffmann, S. Srivastav, and H. K. Genant, ''Quantitative ultrasound of the tibia depends on both cortical density and thickness,'' Osteoporosis Int. **12**, 28–34 (2001).
[17] R. J. Donato, ''Amplitude of p-head waves,'' J. Acoust. Soc. Am. **36**, 19–26 (1964).
[18] D. Royer and E. Dieulesaint, *Elastic Waves in Solids I. Free and Guided Propagation* (Springer-Verlag, Berlin, 1999).
[19] C. Njeh, C. Wu, B. Fan, D. Hans, T. Fuerst, Y. He, and H. K. Genant, ''Estimation of wrist fracture load using phalangeal speed of sound: An *in vitro* study,'' Ultrasound Med. Biol. **26**, 1517–1523 (2000).
[20] M. R. Stegman, R. P. Heaney, D. Travers-Gustafson, and J. Leist, ''Cortical ultrasound velocity as an indicator of bone status,'' Osteoporosis Int. **5**, 349–353 (1995).
[21] A. J. Foldes, A. Rimon, D. D. Keinan, and M. M. Popovtzer, ''Quantitative ultrasound of the tibia: A novel approach for assesment of bone status,'' Bone **17**, 363–367 (1995).
[22] P. Nicholson and M. L. Bouxsein, ''Ultrasonics studies of cancellous bone *in vitro*,'' *Quantitative Ultrasound. Assessment of Osteoporosis and Bone Status* (Martin Dunitz, London, 1999), pp. 195–219.

Depth-dependent acoustic features of diving sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico[a)]

Aaron Thode[b)]
Ocean Engineering Department, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

David K. Mellinger
Cooperative Institute for Marine Resources Studies, Oregon State University, Newport, Oregon 97365

Sarah Stienessen
Alaska Fisheries Science Center, National Marine Fisheries Service, Seattle, Washington 98115

Anthony Martinez
Southeast Fisheries Science Center, National Marine Fisheries Service, Miami, Florida 33149

Keith Mullin
Southeast Fisheries Science Center, National Marine Fisheries Service, Pascagoula, Mississippi 39568

(Received 25 September 2001; revised 25 March 2002; accepted 4 April 2002)

Three-dimensional dive trajectories of three sperm whales in the Gulf of Mexico have been obtained by measuring the relative arrival times and bearings of the animals' acoustic multipath reflections, using two elements of a towed hydrophone array deployed at an unknown depth and orientation. Within the first 6–12 min of the start of a dive, the intervals between successive ''clicks'' of all three whales corresponded closely with the two-way travel time of an acoustic pulse traveling vertically between the animals' position and the ocean bottom. The click spectra contained multiple peaks, including a faint band of energy originally centered near 10 kHz. As the animals descended over 500 m in depth, the center frequency of this band shifted to nearly 15 kHz, but subsequently remained near this value during the rest of the dive. This frequency shift is consistent with that expected from energy scattering from an ensemble of incompressible small-scale air-filled resonators, with diameters on the order of 4 mm. One possible candidate for such an ensemble is proposed to reside in the collapsed frontal sac of the animal. A comparison of the received levels for the bottom and direct multipath arrivals indicates that the whales' acoustic directivity must range between 10–30 dB in the 5–20-kHz region. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1482077]

PACS numbers: 43.80.Ka, 43.30.Sf [WA]

I. INTRODUCTION

In 1957 sperm whales were reported to produce a 6–10-ms pulse called a ''click.'' [1–5] Since that discovery, a considerable body of knowledge has accumulated about these sounds, including statistics on their time-frequency characteristics[6] and repetition rates,[7] possible correlations between animal size and click structure,[8–10] and the identification of regional differences between the acoustic repertoire of separate populations.[11] The function of these sounds remains unknown, although certain acoustic patterns, called ''codas,'' seem associated with social communication.[11,12] The use of these sounds to detect individual prey has been postulated,[4,13–15] and evidence is emerging that sperm whales might echolocate off larger-scale features on the bottom or in the water column.[5,16–19] The mechanisms behind sperm whale sound production remain uncertain,[8,13] despite extensive research into the capabilities of its smaller odontocete relatives.[20]

Here, we present quantitative indications of echolocation behavior from three diving sperm whales in the Gulf of Mexico, and identify depth-dependent features in their vocalization spectra that are consistent with predictions of simple resonator models of small, air-filled swimbladders and/or bubbles.[8,21–24] To obtain these results, multiple acoustic reflections off the ocean surface and bottom[25,26] were exploited to obtain three-dimensional positions of the animals to within a left/right ambiguity. This method is a variant of standard acoustic methods for tracking transient biological sounds.[27–32]

On the evening of 3 July 2000, between 22:00 and 23:30 Central Daylight Time (CDT), a fortuitous set of circumstances converged in the Gulf that allowed three-dimensional acoustic tracking of three diving sperm whales, using only two hydrophones spaced 2 m apart. During this time the National Oceanic and Atmospheric Administration ship GORDON GUNTER was cruising slowly among a group of diving animals, recording their sounds on a hydrophone array deployed at an unknown depth and horizontal tilt behind the ship. Due to very calm ocean conditions on the sea surface and a flat bottom bathymetry, sounds were received at one array element multiple times: from the direct path, and indirect paths reflecting from the surface, the bottom, and both

[a)]Portions of this work have been presented at the 140th (Newport) and 141st (Chicago) Meetings of the Acoustical Society of America, as well as the 14th Biennial Conference on the Biology of Marine Mammals, Vancouver 2001.

[b)]Current address: Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, CA 92093. Electronic mail: thode@mpl.ucsd.edu

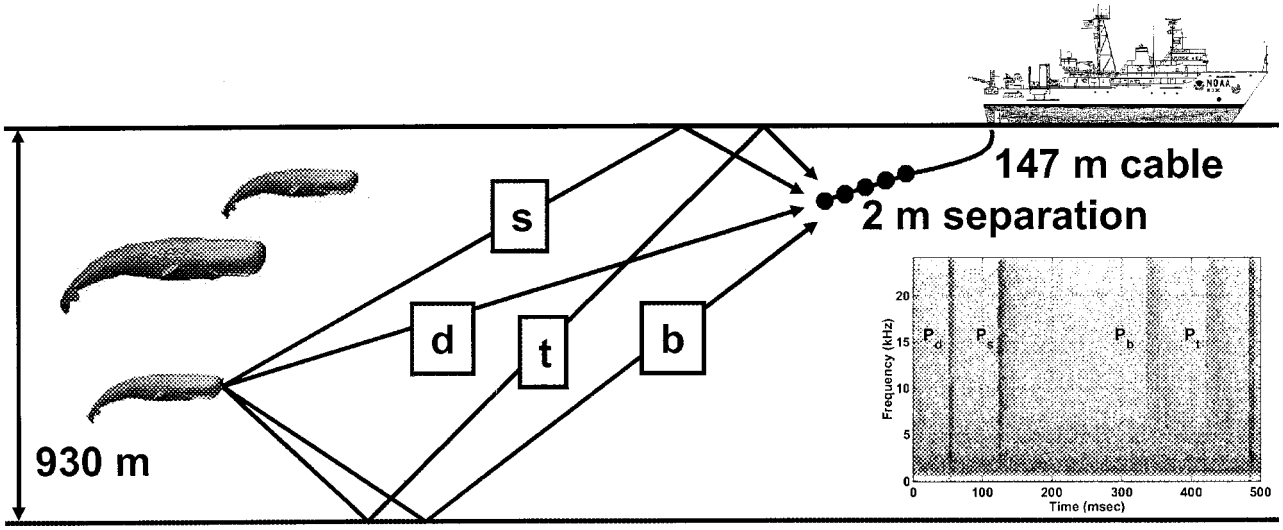


FIG. 1. Illustration of the straight-line ray paths for multiple reflections detected on a five-element array towed by the NOAA ship GORDON GUNTER. On the evening of 3 July 2000, the ship was moving north at 0.5 to 1.5 kn, a speed that allowed the array to sink to depths between 60 and 80 m. Inset: Spectrogram of a direct arrival P_d and associated reflections P_s, P_b, and P_t, which arose from the direct path, surface bounce, bottom bounce, and bottom-surface bounces, respectively. The sampling rate is 48 kHz, the FFT size is 256 points with a Hanning window, and the overlap is 75%. The short FFT length is used to illustrate the three-part temporal structure of the bottom and bottom-surface returns.

the bottom and surface (Fig. 1). The use of a second array element allowed estimation of arrival angles of the direct and surface-reflected paths. Five independent variables could thus be measured for each individual vocalization: three relative arrival times and two array bearings. Thus, both the whale's location (three variables) and the array position (two variables) could be derived simultaneously. The detection of bottom returns effectively enabled the creation of a large-aperture vertical array that could achieve high-resolution depth estimates.

The whales' acoustic behavior in relationship to their depths was investigated. Several authors[5,6,15,17,18] have noted a relationship between click rate and water depth, and a recent paper[16] presents data which suggest that sperm whales may also use sounds to estimate distances to reflective layers within the water column. In this current work, the derived whale depths are compared with their interclick interval (ICI) to investigate possible bottom-ranging behavior by sperm whales during the first 6–12 min of a dive cycle.

A possible correlation between vocalization depth and frequency maxima in the click spectra is also discussed. While changes in depth spectra with dive time have been noted in the literature,[33] the ability to track the animal in depth enables a more detailed analysis. Simple bubble and air-bladder resonator models show a good fit to one observed relationship in the 10-kHz region. The model fit assumes that the resonators involved have dimensions smaller than an acoustic wavelength (on the order of millimeters for a 10-kHz wave), that the resonator volume remains fixed with depth, and that all the resonators have volumes with 30% of the mean. Possible anatomical candidates for these resonators are identified in the frontal sac. If the assumptions behind these models prove valid, then these results suggest that passive detection of acoustic resonances inside vocalizing deep-diving whales is possible.

Finally, the directivity of clicks emitted by sperm whales is studied. This directivity has been the subject of some debate by sperm whale researchers.[14,31] Detecting directivity, or an angular beam pattern, requires the measurement of the same signal at different orientations relative to the whale. This approach is used here: in a situation in which the direct and bottom-reflection paths depart from the whale at significantly different angles, the relative energy of click sounds received on these two paths is measured and compared to an estimate based on a simple acoustic propagation and bottom-reflection model. The comparison provides evidence for a 10–30-dB directivity in the whales' beam patterns.

II. THEORY

A. Estimation of whale range and depth, using reflection arrival times

Whenever three reflections from a click can be isolated in time from the direct path and received on two hydrophones, enough information can exist to determine the whale's range, depth, and azimuth, as well as the array depth and azimuth. Figure 2 illustrates the problem geometry. A whale at range R and depth z_w produces a click that travels a direct path P_d to the array, which is towed at an unknown depth z_a. The direct signal arrives at an angle η_d relative to the forward end of the array (Fig. 2, inset). If the ocean surface is relatively calm, then the signal can also be reflected coherently from the surface, arriving at the array after traveling a distance P_s and arriving from a bearing η_s. If the ocean bottom is flat and its depth H is known, then energy that travels along the path P_b is also received at the array. However, due to scattering from bottom roughness, this reflected energy becomes decorrelated across the array aperture, making arrival angle estimation difficult. Finally, energy may be reflected twice from the ocean boundaries. In the examples presented here, the reflection path P_t ("t" for

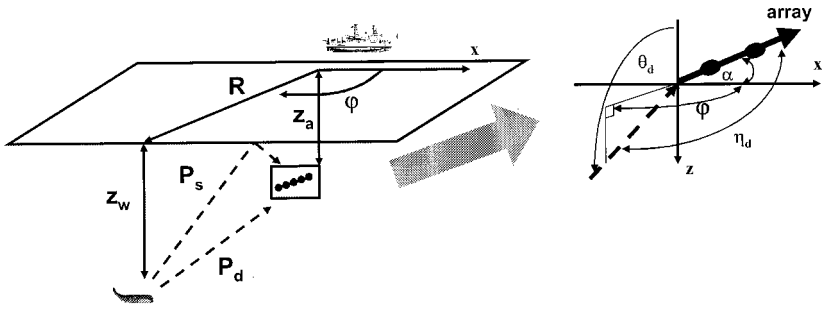


FIG. 2. Geometry of the problem discussed in Sec. II. The array is assumed to lie in the plane of the ship motion (no turns). The dashed lines represent the direct path (P_d) and the surface-reflected path (P_s). Inset: Array angular geometry. The horizontal azimuth of the direct path (dashed line) is ϕ, the angle between the direct path and the vertical is θ_d, and the array tilt from the horizontal x axis is α. The angle between the direct-path arrival along P_d (dashed line) and the array axis is η_d.

twice) involves a bottom-surface reflection, the only double-reflection path that can account for the final arrival in Fig. 1.

An analytical solution for the whale location is presented below. It assumes the ray paths are straight lines, and thus neglects potential effects of sound-speed variation with depth. An analysis of the effects of ray refraction concluded that they were negligible for horizontal source ranges less than a kilometer and source depths greater than 200 m, even under the highly downward-refracting conditions under which the data were collected.

Given that the ray paths are straight lines, the direct and reflected path lengths can be directly related to the whale and sensor locations

$$P_d=\sqrt{R^2+(z_w-z_a)^2},\quad P_s=\sqrt{R^2+(z_w+z_a)^2},$$
$$P_b=\sqrt{R^2+(2H-z_w-z_a)^2},\tag{1}$$
$$P_t=\sqrt{R^2+(2H-z_w+z_a)^2}.$$

Unfortunately, these quantities cannot be measured directly; only relative differences in arrival times between the direct path and each reflection path are available. Assuming the sound speed c is constant with depth and range, the difference between the direct path and reflection path x can be defined

$$\Lambda_x\equiv P_x-P_d=c(t_x-t_d).\tag{2}$$

From (1) and (2), the following relationships are obtained:

$$\Lambda_b^2-\Lambda_s^2\equiv 4H(H-z_a-z_w)+2P_d(\Lambda_s-\Lambda_b),\tag{3a}$$
$$\Lambda_t^2-\Lambda_b^2\equiv 4(2H-z_w)z_a+2P_d(\Lambda_b-\Lambda_t),\tag{3b}$$
$$P_s^2-P_d^2=4z_az_w.\tag{3c}$$

Solving Eqs. (3a) and (3b) for P_d, and setting each result equal to the other yields the following relationship between z_a and z_w:

$$z_a=\frac{\alpha_0+\alpha_1H(H-z_w)}{2H+\alpha_1H-z_w},\tag{4}$$

where

$$\alpha_0=\tfrac{1}{4}(\Lambda_b-\Lambda_t)(\Lambda_s-\Lambda_t),\quad \alpha_1=(\Lambda_b-\Lambda_t)/(\Lambda_s-\Lambda_b).$$

The addition of Eq. (3c) yields a quadratic equation that can be solved for the whale depth z_w

$$0=[\alpha_2H(2+\alpha_1)-(\alpha_0+\alpha_1H^2)\alpha_3H]+z_w[\alpha_1\alpha_3H^2$$
$$-\alpha_2-(\alpha_0+\alpha_1H^2)\alpha_4]+z_w^2H\alpha_1\alpha_4,\tag{5}$$

where

$$\alpha_2=2\Lambda_s(\Lambda_b-\Lambda_t)(\Lambda_s-\Lambda_t-\Lambda_b),$$
$$\alpha_3=16\Lambda_s,\quad \alpha_4=8(\Lambda_b-\Lambda_t-\Lambda_s).$$

Finally, the whale range R can be shown to be

$$R=\sqrt{\frac{[\Lambda_t^2-\Lambda_b^2-4(2H-z_w)z_a]^2}{4(\Lambda_b-\Lambda_t)^2}-(z_a-z_w)^2}.\tag{6}$$

For most situations, only one root from Eq. (5) yields a depth less than H. For source depths near the bottom (typically greater than 900 m for the data presented here), both roots yield depths between the surface and seafloor, but only one yields a possible location when inserted into the expressions for array depth and source range.

A numerical analysis of the effects of timing errors on the localization results is straightforward. Experimentally, the greatest measurement errors arise from the arrival time estimates for the two bottom returns, because the bottom return was usually decorrelated with respect to the direct signal. As a result, the initial bottom arrival times had to be estimated manually, producing estimation uncertainties of up to 0.01 s, or a 480-point uncertainty for a 48-kHz sampling rate. The localization uncertainties generated by a fixed timing error increase with source depth. At the maximum source depth discussed here, 700 m, and for a maximum timing error of 0.01 s, the expected range error is $\pm$400 m, the depth error is $\pm$50 m, and the array depth error is $\pm$20 m, for a receiver depth of 70 m. Generally, the source range estimate has the greatest error percentage, and the source depth estimate has the least. The precision of the depth estimate is not surprising, as the bottom reflections effectively create a vertical array with nearly 1.8-km aperture. Previous literature on acoustic tracking discusses in detail how a large spatial separation between sensors aids localization precision.[27]

B. Recovery of whale azimuth and array tilt

Equations (1)–(6) require the use of only one hydrophone. However, to obtain the azimuth, and thus the full 3D location of the whale, the direct signal must be recorded on

two spatially separated hydrophones. If the towed array tilt is unknown, the bearing from a reflection must also be estimated to recover the two unknown angles, to within a left–right ambiguity.

The angular geometry of the towed array is displayed in the Fig. 2 inset. The x axis is oriented parallel to the ship's velocity, and the positive z axis points toward the ocean floor. The array is assumed to lie in the $x-z$ plane, titled at an angle α relative to the x axis. When energy arrives from azimuth ϕ (measured from the x axis) and elevation θ (measured from the vertical), its bearing relative to the array axis will be measured as η. The relationships between these angles are provided by

$$\cos \eta_x = \cos \theta_x \sin \alpha + \sin \theta_x \cos \alpha \cos \phi, \tag{7}$$

where the index x refers to the array bearing of the direct path (''d'') or surface reflection (''s''). Equation (7) thus yields two equations for the direct (η_d) and surface (η_s) bearings. As horizontal refraction effects are assumed negligible, the horizontal azimuth ϕ is the same for both paths. From the relationships

$$\sin \theta_d = R/P_d, \quad \sin \theta_s = R/P_s,$$

$$\cos \theta_d = (z_\alpha - z_w)/P_d, \quad \cos \theta_s = (z_a + z_w)/P_s,$$

these equations can be inverted, yielding the desired angles

$$\begin{aligned} \sin \alpha &= [P_s \cos \eta_s - P_d \cos \eta_d]/2z_w, \\ \cos \phi &= [P_d \cos \eta_d (z_w + z_a) - (z_w - z_a) P_s \cos \eta_s]/ \\ &\qquad 2Rz_w \cos \alpha. \end{aligned} \tag{8}$$

As α must lie between $\pm 90°$, its value is determined uniquely. The azimuth ϕ can then be resolved to within a left/right ambiguity.

Estimating the arrival angles for the bottom returns was not found to be feasible, as the cross correlation of the diffuse returns between hydrophones yielded multiple ambiguous maxima.

III. METHODS

A. Equipment

The data were recorded on a five-element hydrophone array built by Sonatech, Inc. The elements were ceramic cylinders spaced 2 m apart, with 75 m of trailing cable behind the last element. The individual elements have unamplified omnidirectional sensitivities of 185 to -187 dB *re*: 1 V/μPa between 1.5 and 20 kHz. A preamplifier embedded in each sensor provided 57-dB signal gain, so that the array output sensitivities were effectively -128 to -130 dB *re*: 1 V/μPa. A high-pass single-pole filter with a break frequency of 1 kHz was embedded in the electronics, producing a 6-dB/octave rolloff at low frequencies, in order to reduce cable and ship-induced noise. The array was deployed from a cable 147 m behind the NOAA ship GORDON GUNTER, a former U.S. Navy vessel equipped with vibration-isolated diesel electric engines to minimize acoustic noise. Sixty pounds of weight were added approximately 20 m in front of the first array element to keep the array deep at high tow speeds.

During the time period to be discussed here, the array was being used to track a group of sperm whales in close proximity to the ship. To remain near the animals, the ship had to slow to 0.5–1.5 knots, a speed that caused the array to sink 60–80 m beneath the surface. Over 90 min the ship slowly cruised over the animals while maintaining a fixed heading and speed.

Sound signals captured by the array hydrophones were first routed through a Mackie 1604-VLZ PRO Mic/Line preamp mixer, then passed through an Avens model 4228 8 dual channel variable frequency filter. The latter further reduced the substantial flow noise below 1 kHz. The signals were then recorded onto a TASCAM DA-78HR digital recording system for later analysis. Water depth was measured and recorded by a Simrad EK-500 depth sounder as the ship traveled.

To permit real-time monitoring the TASCAM analog outputs were then entered back into the mixer, where the signal could be monitored by an operator and/or passed through to a Data Translation DT-3809 A/D card inside a Dell Workstation 410. The digitized signals could then be inspected visually via scrolling spectrograms using the proprietary software package ISHMAEL,[34] which also had the capability of calculating bearings to animals in real time.

B. Data analysis procedures

When reflections like those shown in the Fig. 1 inset were either observed on the ISHMAEL display or detected over headphones, the time and water depth were noted and recorded. Back on shore the digital data on the TASCAM tapes were converted directly into WAV-format hard-disk files, using a Spectral Dynamics digital converter. The files were then imported into MATLAB for analysis. A simple click-detection algorithm digitally filtered the signals above 5 kHz and estimated the onset time of clicks and click reflections. Onsets were detected by comparing the variance of a data sample with the mean variance of 50 samples that were known to contain no signal, and classifying the result as a click if the sample variance exceeded the measured noise variance by a factor of 10. The onsets were hand-corrected by visual inspection of the extracted time series. While detecting and adjusting the onset of direct and surface reflections was easily automated using cross-correlation techniques, human judgment was required to select the onset of the diffuse bottom results. The resulting bottom onset timing errors could reach 0.01 s, even when the signals were heavily filtered to emphasize the energy from the bottom, due to the long rise times of the return envelope. Human judgment was also required to classify the arrival as a direct path, surface reflection, etc., a task that required some practice when multiple animals were vocalizing.

The greatest potential difficulty in the analysis was determining which set of bottom/bottom-surface reflections corresponded with a particular direct/surface arrival. At shallow whale depths, the recorded interval between the direct arrival and its associated bottom bounce could in principle exceed the interclick interval (ICI) between direct arrivals. In situations where such confusion was possible, two checks were performed. First, sudden changes in ICI were matched

with corresponding changes in the bottom-reflection intervals. Second, the computations in Sec. II were redone using the interval between the direct arrival and the *second* subsequent bottom and bottom-surface reflections as input data. If the ICI was indeed less than the direct-bottom arrival time difference, then this procedure yielded physically realizable answers. Otherwise, negative or imaginary distances were obtained.

The first two adjacent hydrophone elements were used to estimate bearings, using standard cross-correlation techniques.[35,36] While in theory clicks recorded from the first and fifth element would have provided greater bearing resolution, the surface reflections were found to decorrelate over short spatial distances, making bearing estimation difficult when using sensors separated more than 2 m. The reason behind this difficulty was most likely surface roughness from the relatively calm swell, or even the wake produced by the GUNTER. Once the times and bearings were estimated, the animal and array positions were computed using Eqs. (4)–(6). Unphysical locations were rejected from the analysis, which resulted in rejection rates at times exceeding 25%, whenever the bottom returns were particularly weak.

IV. RESULTS

A. Data description

The data presented here were collected on the evening of 3 July, 2000 between 22:00 and 23:30 CDT. During that period, the GUNTER cruised between 27.8875 N, 89.4823 W and 27.9197 N, 89.4822 W, in water whose depth varied between only 930 to 953 m over tens of square kilometers. The GUNTER maintained a heading of 352 deg in a slight current, at a speed between 0.5 and 1.5 kn. The seas were relatively calm, with a wave height and 1-m swell, both arriving from 70 deg true. The starting times of the three trajectories shown here were 22:23:20 (whale 1), 22:46:40 (whale 2), and 23:15:01 (whale 3).

B. Recovered dive profiles

Figure 3 displays an example of a derived three-dimensional dive trajectory relative to the array, which was moving due north at an average speed of 1 kn, or just under 0.5 m/s. The derived dive times (~11 min) and dive rate (~1 m/s) indicated here match well with previous studies, which used dive tags.[37,38]

Two types of points are shown in subplots (a) and (b): those that provide usable surface reflection bearings, and those that provide only time-of-arrival information. The former allows computation of a full three-dimensional position, whereas the latter yields just range and depth. A third-order polynomial fit has been drawn through each parameter in the form of a solid line. The accompanying dashed line shows the effects of incorporating ray-refraction effects into the computations.[39] Clearly, the error introduced by assuming straight-line ray paths at these short ranges is negligible, a result similar to that of Ref. 32. Much larger localization errors are introduced by measurement uncertainties relating to the bottom arrivals. The localization scatter is consistent with that predicted by the numerical analysis discussed in Sec. II A, given bottom timing errors of 0.01 s or less. The range estimate scatter increases to ±200 m as the animal reaches 700-m depth, which is also consistent with the numerical prediction that the range errors should increase with animal depth.

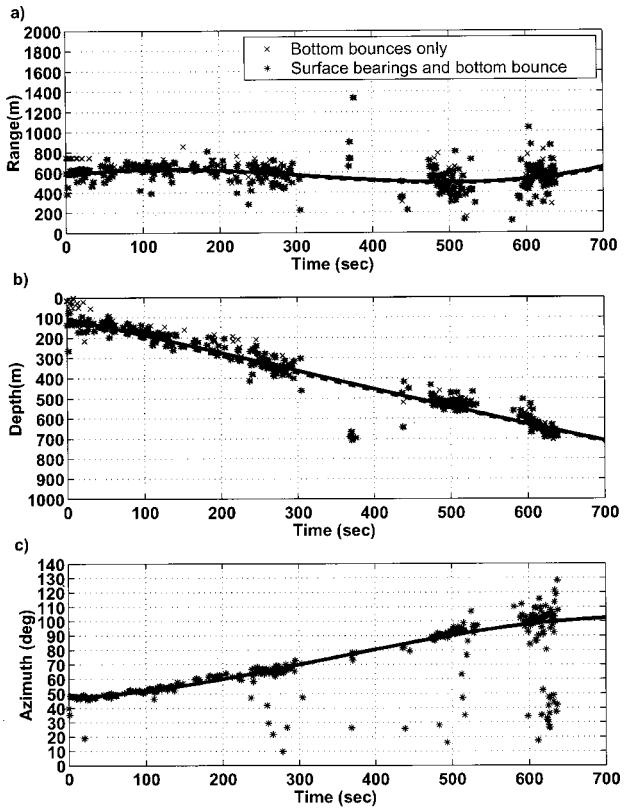


FIG. 3. Three-dimensional trajectory of whale 2, starting at 22:46:40 CDT, measured relative to the moving array, in terms of (a) horizontal range R; (b) depth z_w; and (c) horizontal azimuth ϕ relative to the forward end of the array, which has been assumed to be aligned with the ship's heading. All ranges and bearings are measured relative to the towed array. Crosses represent clicks that yield only relative arrival times, while stars represent clicks that also provide surface-reflected path-bearing information. Solid lines represent a third-order polynomial least-squares fit, and dashed lines represent trajectories corrected for possible sound refraction effects.

Figure 4 shows an example of the array position derived from the dive profile in Fig. 3. The array depth estimates generally lie between 50 and 80 m, consistent with data collected from the time-depth recorder attached to the array during 1–2-kn tows at other times. The derived horizontal array tilt generally lies between 8° and 20°, a fairly large inclination which was probably due to the slow speed of the ship and the weight of a large connector at the end of the array.

Figure 5 incorporates the effects of ship motion to produce the smoothed three-dimensional trajectory estimates of the three whales during the initial descent phase of their dives, relative to the ship location at 22:23:20 CDT. The tracks have been plotted starboard of the ship, although this cannot be verified due to the left/right ambiguity of the measurement. The trajectory marked with an asterisk (''*'') corresponds to the polynomial fit for whale 2, shown in Fig. 3. The range uncertainty is about ±100 m at the top of the dive profiles, and ±200 m at the bottom. To within these uncertainties, it appears that the animals are diving quite

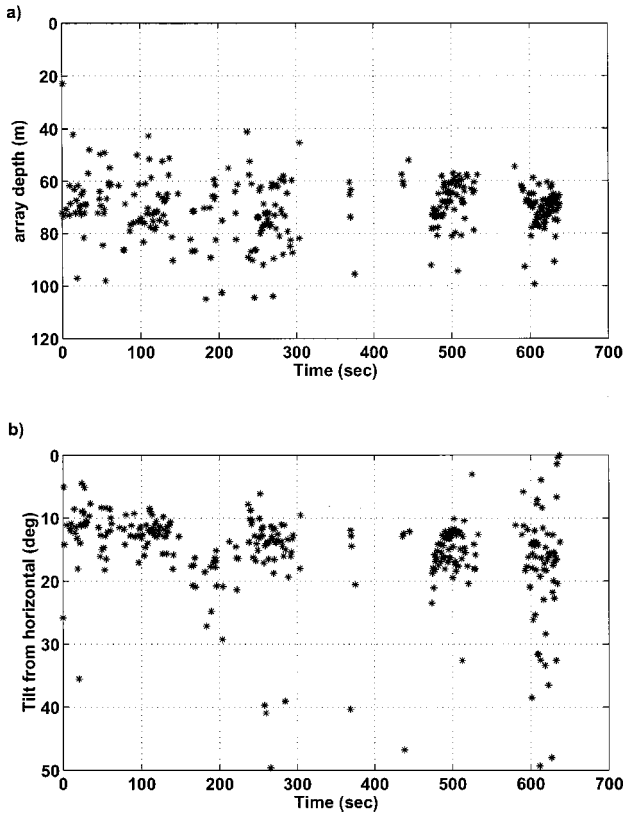


FIG. 4. Example of computed (a) array depth z_a, and (b) array tilt α, extracted from the dive trajectory displayed in Fig. 3.

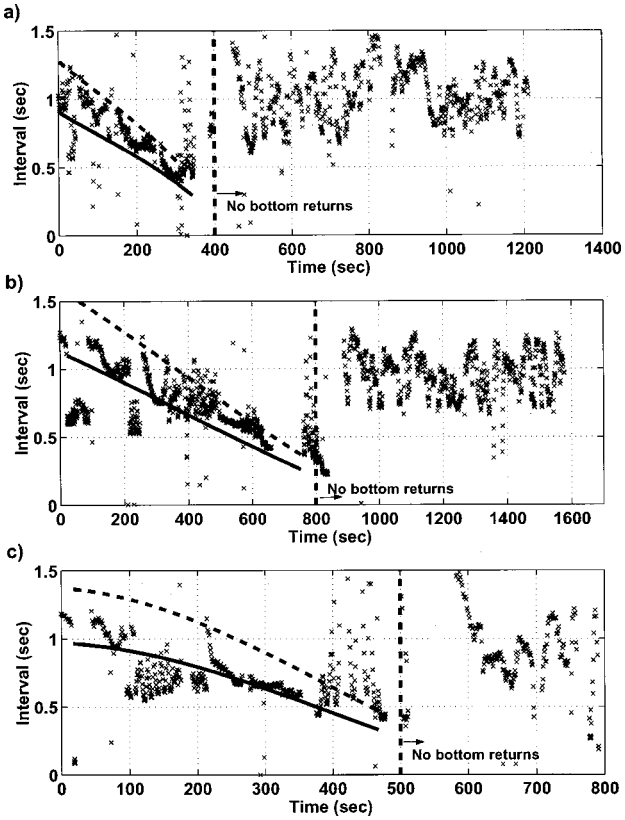


FIG. 6. Relationship between interclick interval (crosses), vertical two-way acoustic travel time between three whales and the ocean floor (solid line), and two-way travel time of a directed pulse along a path 45° with respect to the horizontal (dashed line). Subplots (a)–(c) correspond to the dive profiles of whales 1–3, respectively. The travel times shown in subplot (b) were computed using the depths from whale 2, displayed in Fig. 3. The dashed vertical lines indicate the times after which the bottom bounces disappeared and 3D localization became impossible. Note the sudden change in ICI that occurs after the bottom bounces vanish.

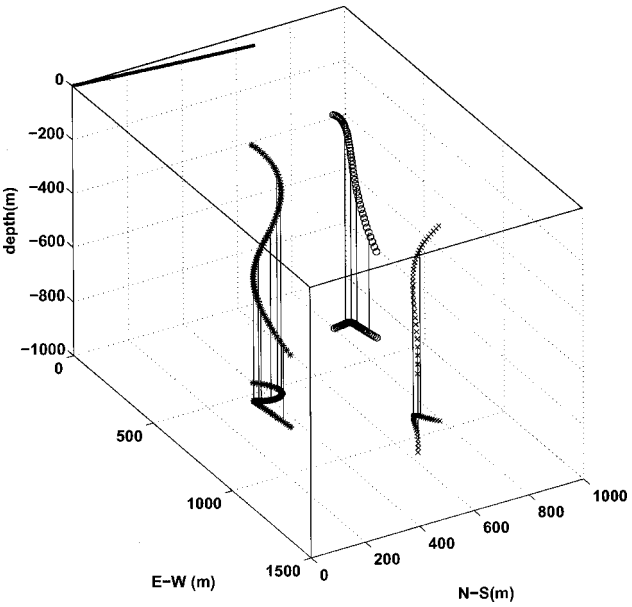


FIG. 5. Smoothed three-dimensional trajectories of whale 1 (“×”), whale 2 (“*”), and whale 3 (“○”), plotted in a stationary reference frame with the origin centered at the ship location at 22:23:20 CDT. The ship track appears as a black horizontal line at the surface. Two-dimensional projections of the tracks in the E–W and N–S plane are also displayed. The tracks have been assumed to be on the starboard side of the ship, although they may be on the port side as well, due to the left/right ambiguity of the localization method.

steeply, at angles ranging between 45° and 90° from the horizontal plane. This observation has implications for the discussion in Sec. V.

C. Relationship between ICI and whale depth

The interval between subsequent clicks, or interclick interval (ICI), has been a topic of some interest in the literature, as outlined in the Introduction. Figure 6 plots the relationship between the ICI and the derived two-way travel time of an acoustic pulse between the animals’ location and the ocean floor, where each subplot corresponds to a different animal. The regions to the left of the vertical dashed lines correspond to times where bottom bounces were often detected, allowing a full three-dimensional fix. The regions to the right correspond to times wherein bottom returns were absent, precluding any precise localization beyond an approximate bearing.

In the region left of the dashed lines, the crosses represent the ICI, and the solid lines represent the two-way travel time of a pulse traveling vertically between the whale and the ocean floor, using the smoothed polynomial depth trajectory fit. The dashed line represents the travel time for a pulse traveling at an angle of 45° from the horizontal, the shallowest angle apparent from the trajectories in Fig. 5. This last

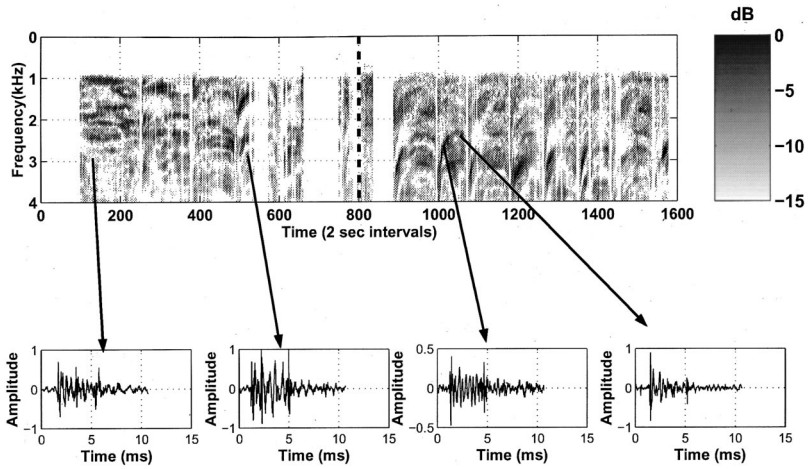


FIG. 7. Averaged spectral content of clicks vs time for whale 2, using the same time scale as Fig. 6(b). The top image shows the uncalibrated click spectral content below 4 kHz. This image is not a continuous spectrogram, but is a sequence of time-averaged click spectra, arranged in 2-s bins of continuous time. Each spectrum was created by extracting a 512 point (11-ms) time segment of the complete click, zero-padding to 2048 points, and then applying an FFT to the zero-padded segment. The long FFT size was used to improve the figure clarity. The four insets show examples of unfiltered click time series used to create the spectra at the indicated times. The time segments illustrate both the pulsed structure of the click and the absence of any contamination from the surface path.

line assumes that the sperm whale signal has a high directivity, and that the maximum value of the whales' beam pattern aligns with the animal's swimming direction. The close correspondence between the vertical two-way travel time and the initial ICI suggests that these whales wait for the bottom return from their previous click before generating the next one, a pattern that has been observed in dolphins as well.[20] However, there are times when the ICI is much smaller than the two-way vertical travel time, like the segment at 220 s in Fig. 6(b), or the segment at 100–200 s in Fig. 6(c).

An interesting transition occurs once the bottom returns vanish. After this point, marked by the vertical dashed line, the ICI increases and follows no discernible pattern, although occasional bouts of rapidly decreasing click interval occur (not to be confused with ''creak'' sounds,[19] which can reach repetition rates of 200 clicks per second). These click segments, followed by silence, are suggestive of hunting behavior.[17,19,40] This behavioral dichotomy between steadily decreasing ICI and irregular ICI has been recently observed in Mediterranean sperm whales as well.[16,41] In light of this, the phase of the dive in which the whale's depth is not increasing, and in which the ICI shows no pronounced relationship with depth, will be labeled *presumed foraging*.

D. Relationship between click spectrum and time

The frequency spectra of sperm whale clicks have been well documented in previous literature (e.g., Refs. 1, 4, 6, 31). Generally, most of the acoustic energy is present at frequencies below 4 kHz. For example, in females the peak regions are typically near 1.2 and 3.0 kHz, respectively,[6] although diffuse energy up to and past 20 kHz has been noted.[3]

It has been noted that the click frequency spectrum from an individual animal changes during the course of a dive.[6,31] Figure 7 illustrates the spectral evolution of clicks collected during the dive sequence illustrated in Figs. 3 and 6(b), for frequencies below 4 kHz. The top image is not a spectrogram; instead, it is similar to a waterfall plot of click spectra, where energy density is shown on a gray scale. The spectrum of each click was computed by conducting a 2048-point FFT on 512 points (11 ms) of Hanning-windowed data. The automatic detection algorithm discussed in Sec. III B initially selected the data segment, but the segment start was then hand-adjusted to begin 10–50 points before the direct path arrival, in order to reduce extraneous noise. As shown in the insets, the 512-point window incorporated the entire click, but excluded any possible contamination from the surface reflection. The evenly spaced peaks visible in the insets are not due to surface reflections, but are part of the click structure. The mechanism behind these ''intraclick'' pulses has been thoroughly explored in previous literature.[2,6,8,9,19]

The 768 points before and after the data segment were zero padded, so the total FFT size was 2048 points. The large FFT length does not extract any additional spectral information from the signal, but it effectively interpolates the spectra to make the features easier to discern in a gray-scale plot. Click spectra were averaged into 2-s continuous time bins before being displayed as shown in Fig. 7.

The time scale shown in the figure is the same as shown in Fig. 6(b), and the dashed vertical line once again indicates the time at which bottom returns vanish and 3D positioning becomes infeasible. There are spectral changes on both short and long time scales. Most obvious are the rapid shifts in peak frequency from around 3.8 to 2.5 kHz over intervals 90 s or less in duration. These rapid frequency sweeps are most visible during the presumed foraging phase (800–1600 s), when the ICI is also changing rapidly. Examples of clicks from the beginning (1010 s) and end (1050 s) of one of these rapid frequency shifts are shown in the insets. Short-term changes like those shown here could be due to both changes in animal behavior and/or orientation; nothing further is deduced from them here.

A longer-term trend is visible between 100 and 700 s, when the animal is known to be increasing depth steadily. Over this time an initial frequency centroid of 2 kHz seems to be shifting upward to nearly 3 kHz. However, this trend is also mixed with complex short-term changes, and simple correlations between animal depth and frequency content cannot be established with certainty.

However, an interesting phenomenon appears if the frequency band of interest is extended to 20 kHz, as shown for all three whales in Fig. 8. A weak, diffuse energy plateau is visible for the entire duration of all three dives. The energy content this band is very weak—10 to 20 dB lower than the energy below 4 kHz. During the animals' initial descents (left of the dashed lines), the frequency centroids of these

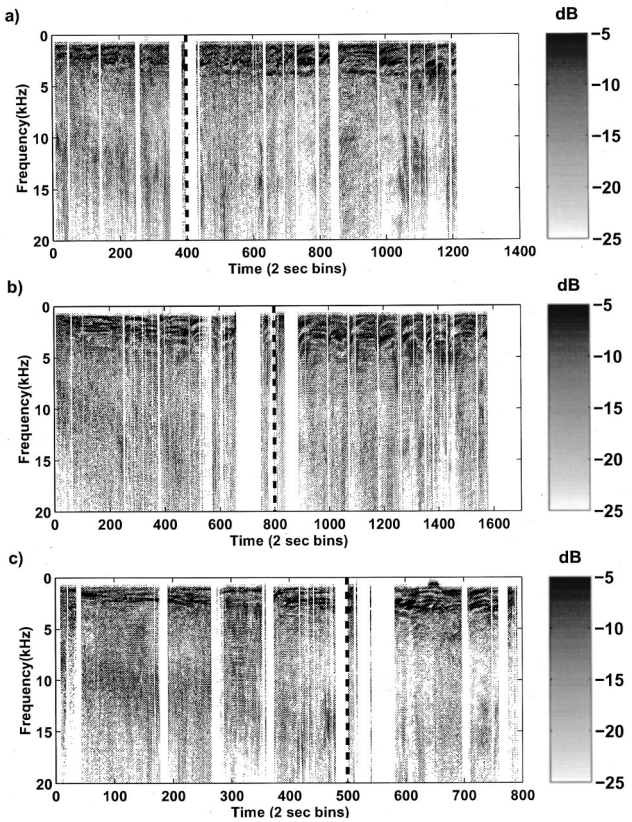


FIG. 8. Click spectral content versus time for frequencies up to 20 kHz for whales 1–3 (a)–(c), respectively. The time scales are identical to those used in Fig. 6. Times to the left of the vertical dashed line represent times when the whale is descending in depth, while times to the right are believed to be associated with constant-depth foraging behavior. Note the presence of a broad, faint band of energy ranging between 10–15 kHz. The intensity scales of these plots have been shifted 5 dB relative to Fig. 7.

bands increase from 10 kHz to about 15 kHz, a 50% shift. Once the animal reaches depth and the bottom bounces vanish, the frequency centroid shifts very little from 15 kHz. The short-term frequency shifts visible in the 0–4-kHz band are not visible here.

An attempt was made to conduct short FFT analyses of 64-point segments within each click, to check whether the intensity of the 10–15-kHz features varied during different portions of the click. Unfortunately, although the band intensity seemed slightly stronger during the first 1.3 ms of the click, no firm conclusion could be established.

V. DISCUSSION

A. Evidence of echolocation off the ocean bottom

The correlation between the animal's depth and interclick interval (ICI) visible in Fig. 6 is a characteristic behavioral pattern of other echolocating animals such as the smaller odontocetes and bats when they are closing on a target.[20,42,43] Possible relationships between sperm-whale ICI and bottom depth have been noted anecdotally,[6,17,19,40,44] and sparse quantitative observations of this phenomenon exist.[5,18] Recent observations by Zimmer *et al.*[16] even suggest sperm whales may also use sounds to estimate distances to reflective layers in the water column.

The data shown here provide additional quantitative support for the bottom-ranging hypothesis, if certain assumptions are made. The first is that the acoustic signal is sufficiently intense that ocean bottom echoes would be audible to the whale. This could happen either if the signal were strong enough omnidirectionally, or if the animal were oriented roughly vertically, so that any potential beam-pattern maximum of the whale[31] is directed toward the ocean floor. The estimated trajectories of Fig. 5 indicate that the animal orientation is indeed close to vertical. The assumption of vertical directivity is also attractive in that it would explain why the bottom returns completely vanish whenever the animal reaches depth and initiates its presumed foraging phase.

A second assumption is that the depth measurements are accurate enough to distinguish between echolocation off the ocean floor, and echolocation off a midwater reflecting layer above the ocean floor. Due to uncertainties in the whales' depth, and the fact that an echolocating animal usually waits a short while between detecting a return and producing the next pulse, a possibility remains that the animals might be detecting a sound-reflecting layer up to 100 m above the ocean floor.

Finally, it has been assumed that the correlation exists because the animal is initially interested in locating the ocean bottom as it dives to foraging depth. However, a curious consequence of this interpretation is that a large fraction of the clicks in Fig. 6 would be dedicated to locating the bottom, although one might expect that only an occasional click would be necessary for the animal to fix its position. For example, Jaquet *et al.*[17,18] have found that male sperm whales off the New Zealand coast have an ICI correlated with bottom depth, but this correlation appears only during the first few seconds of the dive. Does the bottom play a relatively more important role in the foraging strategy of sperm whales in the Gulf of Mexico? A possible alternative interpretation is that the animals find the bottom return a nuisance, and actually time their calls to ensure that a bottom return will not inadvertently overlap a return from a nearer target of greater interest.

B. Depth-dependent interpretation of observed spectral frequency shift

It would be difficult, if not impossible, to isolate depth-dependent effects from other potential behavioral and directional effects, using only spectral information between 0–4 kHz in Fig. 7, due to the complexity of the spectral structure over short-time scales. Additional data on the animal orientation, or acoustic data from sensors spaced a few hundred meters apart would be necessary to identify and isolate these various contributions to the click structure.

However, the long-term frequency centroid shifts between 10–15 kHz visible in Fig. 8 are not confounded by complex short-term effects, so it is reasonable to examine whether changes in animal depth alone can explain the observed shift. Dive-tag data from other experiments have shown that once their initial descent is completed, sperm whales do not make large depth changes until they are ready to surface.[37,38,45] In a similar fashion, the observed frequency shift occurs only when the animal is known to be descending

in depth, and shows no further shifts during the presumed foraging phase. As the observed frequency shifts track the measured and expected depth changes during two distinct phases of the dive cycle, it is reasonable to propose a causal mechanism between the two. This section proposes that depth-dependent resonances in incompressible air-filled cavities a few millimeters in size provide one such mechanism.

1. Discussion of alternative explanations for frequency shift

There are three reasons why the observed frequency shift seems unlikely to be caused by directional effects, or changes in the animals' orientations relative to the hydrophone. First, the observed frequency shifts in the 10-kHz band for all three whales occur only during the descent phase (times left of the vertical black lines). During the presumed foraging phases shown in Fig. 8, the frequency centroids of all three whales shift only slightly from 15 kHz, even though one would expect that the animals' orientations relative to the hydrophones would be changing extensively throughout the dive. Second, the 10-kHz band shows none of the rapid changes in peak frequency clearly visible in the 0–4-kHz band (Fig. 7) over the same time interval. This observation is surprising, because one might expect that energy generated at 10 kHz would display even greater directional properties than energy in the 0–4-kHz band, as the wavelength of 10 kHz is 2–3 times smaller than the peak frequencies visible in Fig. 7. The apparent lack of any correlation between the shifts in peak frequency between 0–4 kHz and the frequency centroid in the 10–15-kHz region suggests that the latter arises from a independent source, with dimensions smaller than a 10-kHz acoustic wavelength. Finally, the same frequency shift is visible for all three animals, even though the positions of the animals relative to the towed hydrophones were different, as can be seen in Fig. 5.

The possibility that the whales systematically adjust their spectral structure above 10 kHz during the initial descent, and then fix the frequency maxima near 15 kHz while presumably foraging, cannot be discounted. However, the energy intensity distributed across the 10–15-kHz band is 10–20 dB less than the energy content in the 0–4-kHz band, and from a sonar engineering perspective it seems suspicious that the animals would manipulate this spectral band without dedicating more acoustic power to the band. Granted, the ability to detect and interpret returns in the 10-kHz region would be expected to be evolutionary advantageous, because in general higher frequencies would aid in the detection and identification of 10-cm objects, such as squid. However, frequencies in this region also suffer higher absorption losses in water,[46] and returns from water-filled targets like squid are expected to be weak. Therefore, if the energy above 10 kHz were truly useful to the animals' foraging success, it is puzzling why their clicks have not dedicated more power to this spectral region. Thus, in this paper the shifts in Fig. 8 will be assumed to be by-products of the sound production mechanism, and not actively manipulated by the animal. This so-called ''generator noise'' assumption does not preclude the possibility that this faint band may serve a communicative purpose, just as the low-frequency omnidirectional ''air-borne components'' of other odontocete clicks may serve a useful communicative purpose.[47]

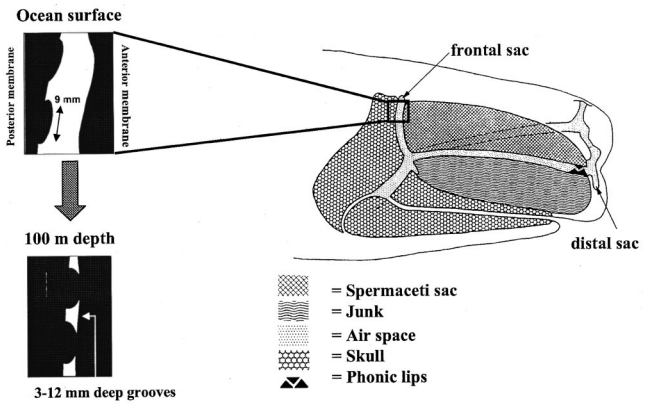


FIG. 9. Schematic of the internal of the sperm-whale head, showing the location of the frontal and distal air sacs. Top inset: Magnified view of frontal sac structure (Ref. 8). The posterior (back) face of the sac is covered by a pavement of rounded, fluid-filled knobs, while the anterior (front) face is smooth. Bottom inset: As the external water pressure increases, the sac walls are expected to press together, trapping packets of interconnected air passages between the knobs. (Main figure first published in Ref. 52, reproduced with permission of Ted Cranford.)

Before restricting subsequent discussion to air cavities, the possibility of depth-dependent effects on the acoustic properties of oil-filled cavities should also be considered. The compressional velocity of spermaceti oil does vary with pressure and temperature.[48] However, careful measurements have found that this speed varies by only 11% over extremes in pressure and temperature, and only 2% if temperature is constant.[48] Therefore, the properties of spermaceti oil could account for at most 20% of the total frequency shift observed, unless a formula with a strong nonlinear dependence on the sound speed of the cavity material can be identified, or unless the animal can generate enormous internal pressures via muscular contraction.

2. Anatomical evidence for air-filled cavities smaller than an acoustic wavelength

After decades of research, there is a general consensus that odontocete sounds are produced in the nasal region by a pneumatic mechanism.[49–52] However, the particular roles played by various air-filled nasal passages and cavities in the sperm-whale sound production mechanism remain uncertain (Fig. 9), although anatomical comparisons with smaller odontocetes suggest that the *museau de singe*,[6,8,13] or ''phonic lips'' at the anterior extreme of the whale's nose are a probable sound production mechanism.[13,52]

Any air cavities present in the sperm-whale head will compress or collapse shortly after a dive starts. For example, the lungs and trachea of most whales are generally believed to collapse by 100-m depth, forcing air into nasal and other air cavities in the head. Indeed, a complete collapse of the respiratory system has been documented at 300-m depth for a bottlenose dolphin.[53] If the internal air cavities do not completely collapse, then the air pressure inside them would increase with the exterior water pressure. This increase may arise by either contraction of the air-cavity dimensions, by

the transfer of air from other contracting or collapsing passages in the respiratory system, or by a combination of the two.

Anatomical studies of the sperm-whale nose have demonstrated the presence of at least two large air sacs, one located anterior and the other posterior of the spermaceti organ, which have been appropriately named the distal and frontal sacs, respectively[8,13,52,54] (Fig. 9). A dissection of the frontal sac revealed the following structure (Fig. 9, top inset):

> "The posterior wall of the sac is covered with a pavement of smooth, rounded, fluid-filled knobs... These knobs vary in diameter from 4 to 13 mm with a mean diameter of 9 mm... Between the knobs lie grooves varying from 3 to 12 mm deep. Scattered throughout these channels are thin membranous transverse septa that divide the channels into many discrete parts..." [8]

By contrast, the anterior wall of the frontal sac is very smooth, so that when the two faces of the sac press together (Fig. 9, bottom inset), a network of interconnected air cavities forms, with cavity dimensions much smaller than the acoustic wavelength of a 10-kHz sound in water (15 cm) or air (34 mm).

Norris and Harvey, who made the above observations, speculated that this interconnected mesh of air helps reflect acoustic energy generated by the phonic lips, and that subsequent reflections between the distal and frontal sacs explain certain features of the sperm-whale click structure.[2,6,9] Even an air film only 1-mm thick will act as a perfect acoustic reflector at frequencies above 3 kHz, and will still reflect 20% of the incident acoustic energy at 1 kHz, assuming that the opposite faces of the sac are not driven together by the acoustic pulse. Therefore, if the Norris and Harvey theory is true, the collapsed frontal and distal sacs play prominent roles in the sound production mechanism, and the air filigree trapped inside the frontal sac is exposed to high acoustic intensities.

3. A simple depth-dependent resonator model for frequency shifts

If small air cavities are indeed excited by the passage or reflection of an incident acoustic pulse, they will have a tendency to radiate energy at a resonant frequency that is determined by a balance between the radiation impedance of the cavity, and the acoustic compliance of the air inside the cavity. If a typical cavity dimension a is smaller than an acoustic wavelength λ (that is, $2\pi a/\lambda \ll 1$), and its boundaries are not infinitely rigid, then the energy radiated by the cavity will tend to be omnidirectional, regardless of the cavity shape.[55,56] In addition, whenever $2\pi a/\lambda \ll 1$, the cavity's fundamental resonant frequency will be determined primarily by the cavity volume, and the cavity can be effectively modeled as a sphere (or ellipsoid), with a being the sphere radius. For example, even an extremely flattened spheroid with a ratio of major-to-minor axis of 16 will have a frequency only 30% higher than a sphere of equivalent volume.[23,57] The limited data from Norris and Harvey[8] suggest major–minor axis ratios of around 3, which would generate resonant frequencies only 5% greater than that of a sphere of equivalent volume. Thus, if the 10-kHz band is assumed to be the product of an ensemble of small-scale resonators of similar volumes, then a very simple spherical resonator model can be reasonably tested against the observations.

The fundamental resonance frequency for a spherical air cavity with $2\pi a/\lambda \ll 1$ is provided by the following two expressions.[21–24] The first expression assumes the cavity volume remains fixed with increasing water pressure, which requires that the cavity be connected, at least indirectly, to another compressible cavity

$$f_{\text{incompressible}} = \frac{1}{2\pi a}\sqrt{\frac{2\gamma P_{\text{water}} + X}{\rho_{\text{water}}}}. \tag{9a}$$

Here, a is the cavity radius, γ is the ratio of specific heats for an ideal gas, P_{water} is the external water pressure, and ρ_{water} is the water density. The quantity X depends on the characteristics of material surrounding the cavity. If the material is water, Eq. (9a) represents a freely oscillating bubble, and $X = 0$. If the material is an infinite elastic solid (the Andreeva model[24]), $X = 4\mu$, where μ is the real part of the shear modulus. A widely accepted model of a fish swimbladder (the Love model[23]) assumes that the air cavity is surrounded by a thin membrane capable of sustaining surface tension, embedded in a viscous fluid. The derivation yields a result of $X = 2s(\gamma - 1)/a$, where s is the surface tension of the tissue at the air/shell interface.

The second expression assumes that the cavity is completely isolated from other air sources, so the cavity dimensions compress under increased pressure. If the original cavity radius is a_0 at water pressure P_0, then from Boyle's law the appropriate expression is

$$f_{\text{compressible}} = \frac{P_0^{-1/3}}{2\pi a_0}\sqrt{\frac{3\gamma}{\rho_{\text{water}}}}P_{\text{water}}^{5/6}. \tag{9b}$$

The best-fit predictions for both incompressible and compressible models are plotted over the observed data in Fig. 10, where the fitted depth trajectory curves have been used to average the click spectra into depth bins. All curves were fit to the 10-kHz band, around a depth of 250 m, the shallowest depth for which data were available for all three animals. The assumption of a compressible cavity clearly does not match the visible data trend, but the incompressible model shows a fairly good fit. A nearly exact fit is achieved by applying the Andreeva and Love models to the incompressible model, by using an additional data point at 650-m depth to fix the extra free parameter. Table I lists the values of a and μ or s that are required to achieve these fits. Generally, the values obtained for the 10-kHz signal seem realistic, in that a shear rigidity modulus of 2×10^6 Pa is the consistency of hard rubber,[58] and surface tensions up to 3.6×10^6 dyne/cm^2 have been measured in the air bladders of certain species of fish.[23]

Fitting the model to the data required the use of incompressible air-filled cavities 2–4 mm in diameter. The previous discussion on sperm-whale anatomy suggests that nu-

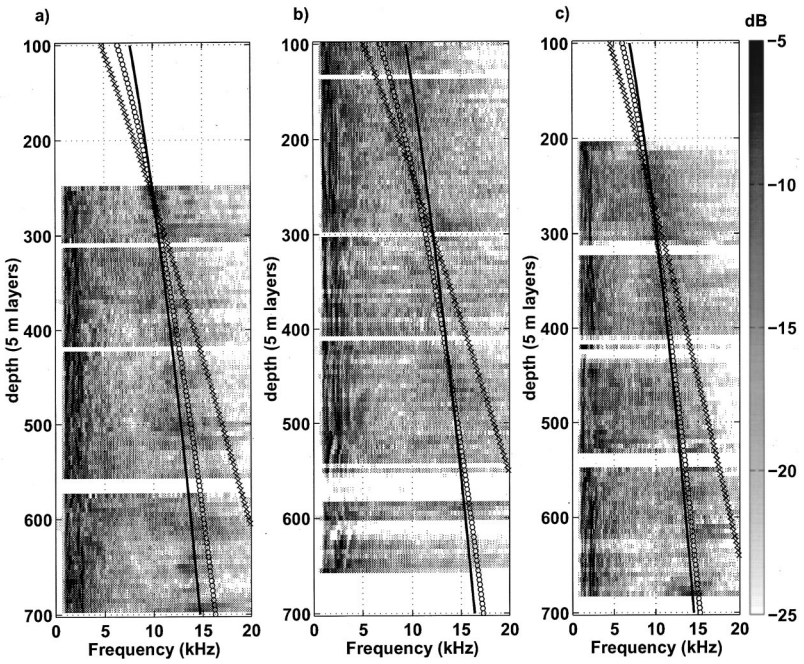


FIG. 10. Comparison of click spectra versus animal depth for whales 1–3 (a)–(c), respectively, for the 10-kHz band. Spectra have been depth-averaged over 5-m intervals. The circles ("○") represent predictions of the incompressible model in Eq. (9a) (with $X=0$). Crosses ("×") represent predictions of the simple compressible bubble model described in Eq. (9b). The solid lines display the predictions of the fitted elastic and viscous swimbladder models.

merous interconnected small passages may exist in the grooves between the knobs on the posterior face of the collapsed frontal sac. Whether these passages remain fixed size with increasing water depth, due to additional air inputs from other air passages, is unknown but plausible, as the cavities are interconnected to each other and the rest of the respiratory tract. Other small-scale air cavities may also exist in the distal sac and surrounding the *museau de singe*.

The greatest weakness of the model is that it assumes all of the proposed resonators have volumes within ±30% of the mean volumes derived in Table I, assuming that the proposed resonances lie within ±1 kHz of the centroid frequency. This is a stringent requirement, and further microanatomical information is required to judge whether such a uniform distribution of small-air cavities is unrealistic.

C. Some comments on signal directivity

The circumstances under which bottom returns are detected are intriguing. As mentioned in Sec. IV C, the returns are always detected when the animal is descending toward the ocean floor, but are never detected once the animal initiates presumed foraging behavior, even though surface re

TABLE I. Parameters used to fit 10-kHz curves in Fig. 10. "Bubble radius" is the radius obtained by assuming the air cavity is surrounded by water. "Elastic bubble radius" refers to radius obtained using Andreeva (Ref. 24) or Love (Ref. 23) model.

	Bubble radius (Refs. 21, 22) (mm)	Elastic bubble radius (mm)	Andreeva (Ref. 24) shear rigidity modulus μ (MPa)	Love (Ref. 23) Surface tension s (10^6 dyne/cm)
Whale 1	1.7	2.0	4.5	0.5
Whale 2	1.5	2.0	1.7	2.2
Whale 3	1.7	2.0	1.0	1.2

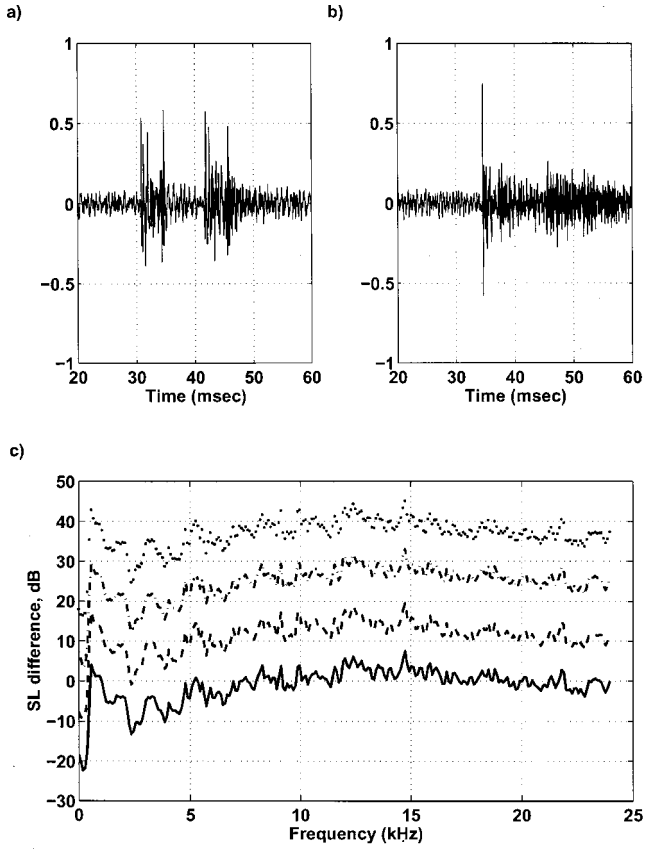


FIG. 11. Argument against omnidirectional signal hypothesis. (a) Direct and surface paths recorded at 128 s, when the second whale was at ~120 m depth. (b) Associated bottom return. (c) Estimated source-level differences between bottom and direct energy arrivals, averaged for 31 clicks with strong bottom returns. Solid line: perfectly reflecting bottom, coherent bottom reflection. Dashed line: incoherent bottom scatter, no bottom loss. Dash-dot line: coherent bottom reflection, assuming a bottom composed of sand (bottom speed 1700 m/s, density 1.0 g/cc). Dotted line: incoherent bottom scatter, energy loss into sand bottom.

flections are clearly present throughout the sequence. Even during the initial descent phase, the detection of bottom returns is sporadic. After long periods during which only direct and surface-reflection paths are recorded, the bottom returns often fade in within seconds, with a 10-dB increase in signal energy that is typically accompanied by energy variation in the direct signal arrival of less than 3 dB. These qualitative observations suggest that the sperm-whale signals have directional properties (see also Ref. 31), also defined as ''directivity.'' [46]

While the 3D localization resolution presented here is too coarse to estimate the whales' instantaneous orientations, and thus their potential directivities, enough information exists to argue that some of the recorded clicks could not have been generated by an omnidirectional source.

By using the sonar equation and crude bottom reverberation models,[46] the original source levels for energy arriving along the bottom and direct propagation paths can be estimated. If the derived source levels are different between the two paths, then the signal cannot be omnidirectional. Various models of the ocean bottom can be invoked to account for the transmission loss into the ocean bottom, ranging from a perfect coherent reflector, to a lossy incoherent scatterer. The latter model best fits empirical measurements of data collected over the past 40 years from both the sonar[46] and radar[36,59] communities.

The following data analysis is based on 31 clicks recorded during the dive profile shown in Fig. 3, during times between 127–130, 375–380, 552–590, and 730–737 s. During these times the bottom return intensities increased markedly over an interval of a few seconds. If the omnidirectional hypothesis is correct, the estimated bottom-path source levels should be equal to or lower than the direct path source level estimate, if one uses equal-time data segments from the *beginning* of both the direct and bottom path arrivals. An accurate analysis thus requires direct and bottom returns with steep rise times, as shown in Figs. 11(a) and (b) for a signal recorded at 128 s from a whale roughly 800 m above the ocean floor. Energy was measured during the first $\tau = 6$ ms of both the direct and bottom-reflected arrivals of the 31 clicks.

The mean differences between these source levels are plotted as a function of frequency and bottom type [Fig. 11(c)]. The solid and dashed curves show the source-level differences assuming no energy loss into the bottom. The dash-dot and dotted curves incorporate a reflection coefficient assuming a sandy-silty bottom half-space, with a density equal to that of seawater and a bottom speed of 1700 m/s.[60] The results support the concept of an omnidirectional source only in the extreme limit of a perfect coherent reflector, which is highly unlikely to represent the ocean bottom in the Gulf. Generally, in the 5–20-kHz region, a 10–30-dB difference in source level must exist between the direct and bottom paths, with the bottom-reflected path containing the larger source level. The true directivity probably lies between these extremes. The precise shape of Fig. 11(c) is not informative, because the reflection coefficient from the ocean bottom is definitely frequency dependent.

Regardless of the exact bottom composition, the bottom-path arrival most likely had a greater initial source level than the direct-path arrival. This analysis thus supports the conclusion of sperm whale directivity by Whitney[29] and Mohl *et al.*[31]

VI. CONCLUSION

Three-dimensional localization of sperm whales using multiple reflections from the ocean surface and bottom, without any knowledge of the array position, has been demonstrated for three animals in the Gulf of Mexico. While the algorithm presented here works well, its long-term practicality is restricted for two reasons. First, the data must be collected in a relatively shallow region with flat bathymetry, and the surface conditions must be calm enough to ensure coherence between surface reflections recorded on two hydrophones. Second, the procedures listed here are difficult to automate, because the classification of the various returns is a complex process that currently requires the experienced judgment of a human, particularly when multiple animals are calling. Consequently, the localization procedure is slow.

Improvements on the methods used here can be achieved by deploying a second towed array a few hundred meters behind the first array, to remove the need for bottom reflections for localization. In addition, depth and orientation information from a tagged animal can be used in conjunction with a towed array to provide much more detailed information on the three-dimensional trajectory and relative orientation of the animal.[16,45] Both approaches are currently being pursued.

While the particular algorithm described in Sec. II works only under specialized ocean conditions, the conclusions reached from their use may have broader applications. Two aspects of the sperm-whale click, the interclick interval (ICI) and the frequency spectrum, appear to contain depth-dependent features. The interpretation of the observed frequency shifts as depth-dependent resonances is intriguing, but validation of the assumptions used in the model requires further anatomical knowledge of the locations and size distributions of small-scale air passages in the sperm-whale head. This issue is important because recent debate on the physical effects of anthropogenic sounds on marine mammals has assigned particular prominence to potential resonant effects inside the animal.[61] The present work suggests that internal resonances may be at least weakly excited by an animal's own vocalizations, and might thus be investigated by passive acoustic methods.

Additional acoustic data collection from different geographical locations, using different array deployments, will be required to solidify the conclusions presented here. What seems certain, though, is that the relatively shallow, flat bathymetry of the Gulf of Mexico is an excellent acoustic laboratory for gleaning insight into the acoustic lives of these mammals.

ACKNOWLEDGMENTS

The authors wish to thank the crew of NOAA ship GORDON GUNTER, whose expertise is exceeded only by their enthusiasm for trying something new. We thank Cadet Andrea Stieger, Peter Tyack, and Jonathan Gordon for assistance in

monitoring and tracking groups overnight. Peter Tyack, Rebecca Thomas, and Patrick Miller also generously allowed access to their transcription and recording equipment at the Woods Hole Oceanographic Institution. Walter Zimmer at the SACLANT center generously shared a draft of a paper on the recent successful SACLANT SIRENA cruise. Jonathan Gordon, Bill Watkins, Ted Cranford, and two anonymous reviewers provided helpful comments on the manuscript, and Ted Cranford provided permission to use Fig. 9. Funding for the sperm-whale research pilot study was provided through an Interagency Agreement (15958) between the U.S. Department of the Interior Minerals Management Service and the National Marine Fisheries Service, with additional support provided by the Office of Naval Research (ONR). ONR provided support for the first and second [ONR Contract No. N00014-00-F-0395] authors of this paper.

[1] L. V. Worthington and W. E. Schevill, ''Underwater sounds heard from sperm whales,'' Nature (London) **180**, 291–293 (1957).
[2] J. H. Backus and W. E. Schevill, ''Physeter Clicks,'' in *Whales, Dolphins, and Porpoises*, edited by K. S. Norris (University Press, Berkeley, 1966), pp. 510–527.
[3] W. A. Watkins, ''Acoustic behaviors of sperm whales,'' Oceanus **20**, 50–58 (1977).
[4] W. A. Watkins, ''Acoustic and the Behavior of Sperm Whales,'' in *Animal Sonar Systems*, edited by R.-G. Busnel and J. F. Fish (Plenum, New York, 1980), pp. 283–290.
[5] J. Gordon, ''The acoustic world of sperm whales,'' Eur. Res. Cetaceans **9**, 29–33 (1995).
[6] J. C. Goold and S. E. Jones, ''Time and frequency domain characteristics of sperm whale clicks,'' J. Acoust. Soc. Am. **98**, 1279–1291 (1995).
[7] H. Whitehead and L. Weilgart, ''Click rates from sperm whales,'' J. Acoust. Soc. Am. **87**, 1799–1806 (1990).
[8] K. S. Norris and G. W. Harvey, ''A Theory for the Function of the Spermaceti Organ of the Sperm Whale (*Physeter Catodon* L.),'' in *Animal Orientation and Navigation*, edited by S. R. Galler, K. Schmidt-Koeing, G. J. Jacobs, and R. E. Belleville (NASA Scientific and Technical Office, Washington, D.C., 1972).
[9] J. C. D. Gordon, ''Evaluation of a method for determining the length of sperm whales (*Physeter catodon*) from their vocalizations,'' J. Zool. **224**, 301–314 (1991).
[10] J. C. Goold, ''Signal processing techniques for acoustic measurement of sperm whale body lengths,'' J. Acoust. Soc. Am. **100**, 3431–3441 (1996).
[11] L. Weilgart and H. Whitehead, ''Coda communication by sperm whales off the Galapagos Islands,'' Can. J. Zool. **71**, 744–752 (1993).
[12] W. A. Watkins and W. E. Schevill, ''Sperm whale codas,'' J. Acoust. Soc. Am. **62**, 1485–1490 (1977).
[13] T. W. Cranford, ''The sperm whale's nose: Sexual selection on a grand scale?,'' Marine Mammal Sci. **15**, 1133–1157 (1999).
[14] W. A. Watkins and M. A. Daher, ''Variable Spectra and Nondirectional Character of Clicks From Near-Surface Sperm Whales (*Physeter Catodon*),'' in *Advances in the Study of Echolocation of Bats and Dolphins*, edited by J. Thomas, C. Moss, and M. Vater (Chicago University Press, Chicago, 2002).
[15] P. L. Tyack, ''Studying How Cetaceans Use Sound to Explore Their Environment,'' in *Perspectives in Ethology*, edited by D. H. Owings, M. D. Beehcer, and N. S. Thompson (Plenum, New York, 1997), Vol. 12, pp. 251–297.
[16] W. Zimmer, M. P. Johnson, A. D'Amico, and P. L. Tyack, ''Combining data from a multisensor tag and passive sonar to determine the diving behavior of a sperm whale (*Physeter macrocephalus*),'' J. Ocean Eng. (to be published).
[17] N. Jaquet, S. Dawson, and L. Douglas, ''Vocal behavior of male sperm whales: Why do they click?,'' J. Acoust. Soc. Am. **109**, 2254–2259 (2001).
[18] L. Douglas, ''Click counting: An acoustic censusing method for estimating sperm whale abundance,'' M.Sc. thesis, University of Otago, Dunedin, New Zealand, 2000.
[19] J. C. Gordon, ''The behavior and ecology of sperm whales off Sri Lanka,'' Ph.D. dissertation, University of Cambridge, 1987.
[20] Whitlow W. L. Au, *The Sonar of Dolphins* (Springer Verlag, New York, 1995).
[21] M. Minnaert, ''On musical air-bubbles and the sounds of running water,'' Philos. Mag. **16**, 235–248 (1933).
[22] N. H. Fletcher, *Acoustic Systems in Biology* (Oxford University Press, New York, 1992).
[23] R. H. Love, ''Resonant acoustic scattering by swimbladder-bearing fish,'' J. Acoust. Soc. Am. **64**, 571–579 (1978).
[24] L. B. Andreeva, ''Scattering of sound by air bladders of fish in deep sound scattering ocean layers,'' Sov. Phys. Acoust. **10**, 17–20 (1964).
[25] D. H. Cato, ''Simple methods of estimating source levels and locations of marine animal sounds,'' J. Acoust. Soc. Am. **104**, 1–12 (1998).
[26] R. Aubauer, M. O. Lammers, and W. W. L. Au, ''One-hydrophone method of estimating distance and depth of phonating dolphins in shallow water,'' J. Acoust. Soc. Am. **107**, 2744–2749 (2000).
[27] W. A. Watkins and W. E. Schevill, ''Sound source location by arrival times on a non-rigid three-dimensional hydrophone array,'' Deep-Sea Res. **19**, 691–706 (1972).
[28] J. L. Spiesberger and K. M. Fristrup, ''Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography,'' Am. Nat. **135**, 107–153 (1990).
[29] W. Whitney, ''Observations of Sperm Whale Sounds from Great Depths,'' Marine Physical Laboratory, Scripps Institute Oceanography Report No. MPL-U-11/68, 1968.
[30] D. E. McGehee, ''1997 Sperm Whale Abundance and Population Structure Cruise Leg II Sonobuoy Project, Final Report,'' Tracor Applied Sciences Report No. T-97-56-0002-U, 1997.
[31] B. Mohl, M. Wahlberg, P. Madsen, L. Miller, and A. Surlykke, ''Sperm whale clicks: Directionality and source level revisited,'' J. Acoust. Soc. Am. **107**, 638–648 (2000).
[32] M. Wahlberg, B. Mohl, and P. T. Madsen, ''Estimating source position accuracy of a large-aperture hydrophone array for bioacoustics,'' J. Acoust. Soc. Am. **109**, 397–406 (2001).
[33] William Watkins and Mary Ann Daher, *Features of Sperm Whale Click Transmission* (Proc. Biological Sonar Conference, Carvoeiro, Portugal, 1998).
[34] D. K. Mellinger, ISHMAEL 1.0 User's Guide, NOAA/PMEL, 2002.
[35] C. W. Clark, ''A real-time direction-finding device for determining the bearing to underwater sounds of Southern Right Whales, *Eubalaena australias*,'' J. Acoust. Soc. Am. **68**, 508–511 (1980).
[36] M. I. Skolnik, *Introduction to Radar Systems* (McGraw-Hill, New York, 1962).
[37] W. A. Watkins, M. A. Daher, N. A. DiMarzio, A. Samuels, D. Wartzok, K. M. Fristrup, P. W. Howey, and R. R. Maiefski, ''Sperm whale dive behavior from radio tag telemetry,'' Marine Mammal Sci. **18**, 55–68 (2002).
[38] V. Papastavrou, S. C. Smith, and H. Whitehead, ''Diving behavior of the sperm whale *Physter macrocephalus*, off the Galapagos Islands,'' Can. J. Zool. **67**, 839–846 (1989).
[39] G. Frisk, *Ocean and Seabed Acoustics* (Prentice-Hall, Englewood Cliffs, NJ, 1994).
[40] J. C. Gordon and R. Leaper, ''Effects of whale watching vessels on the surface and under-water acoustic behavior of sperm whales off Kaikoura, New Zealand,'' NZ Dep. Conserv. Science & Research Series 52, 64–74 (1992).
[41] V. Teloni, W. Zimmer, C. Fossati, M. Manghi, G. Pavan, and M. Priano, ''Variability of temporal and spectral click characteristics of sperm whales (*Physeter macrocephalus*),'' 14th Annual Conference European Cetacean Society, Cork, Ireland, 4-2-2000.
[42] E. K. V. Kalko, ''Insect pursuit, prey capture and echolocation in pipistrelle bats (*Microchiroptera*),'' Anim. Behav. **50**, 861–880 (1995).
[43] J. W. Bradbury and S. L. Vehrencamp, *Principles of Animal Communication* (Sinauer, Sunderland, MA, 1998).
[44] J. C. Gordon and D. Gillespie, *Playback of low power ATOC-like sound to sperm whales* (World Marine Mammal Science Conference, Monaco, 1998).
[45] M. Johnson, P. Tyack, and D. Nowacek, ''A digital recording tag for measuring the response of marine mammals to sound,'' J. Acoust. Soc. Am. **108**, 2582–2583 (2000).
[46] R. J. Urick, *Principles of Underwater Sound*, 3rd ed. (Peninsula, Los Altos, CA, 1983).
[47] T. W. Cranford, ''In Search of Impulse Sound Sources in Odontocetes,'' in *Hearing in Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York, 2000), pp. 109–156.
[48] J. C. Goold, J. D. Bennell, and S. E. Jones, ''Sound velocity measure-

ments in spermaceti oil under the combined influences of temperature and pressure,'' Deep-Sea Res., Part I **43**, 961–969 (1996).
[49] K. S. Norris and G. W. Harvey, ''Sound transmission in the porpoise head,'' J. Acoust. Soc. Am. **56**, 659–664 (1974).
[50] K. S. Norris, ''Cetacean Biosonar: Part 1—Anatomical and Behavioral Studies,'' in *Biochemical and Biophysical Perspectives in Marine Biology*, edited by D. C. Malins and J. R. Sargent (Academic, New York, 1975), pp. 215–234.
[51] S. H. Ridgway, D. A. Carder, R. F. Green, and S. L. L. Gaunt, ''Electromyographic and Pressure Events in the Nasolaryngeal System of Dolphins During Sound Production,'' in *Animal Sonar Systems*, edited by R.-G. Busnel and J. F. Fish (Plenum, New York, 1980), pp. 239–250.
[52] T. W. Cranford, M. Amundin, and K. S. Norris, ''Functional morphology and homology in the odontocete nasal complex: Implications for sound generation,'' J. Morphol. **228**, 223–285 (1996).
[53] S. H. Ridgway and B. L. Scronce, ''Respiration and deep diving in the bottlenose porpoise,'' Science **166**, 1651–1653 (1969).
[54] M. G. P. and H. Beauregard, ''Sur 'l'organe des spermaceti','' C. R. Soc. Biol. **11**, 343–344 (1885).
[55] G. G. Harris, ''Considerations on the Physics of Sound Production by Fishes,'' in *Marine Bio-acoustics*, edited by W. N. Tavolga (Macmilan, New York, 1964), pp. 233–252.
[56] Earl G. Williams, *Fourier Acoustics* (Academic, London, 1999).
[57] M. Strasberg, ''The pulsation frequency of nonspherical gas bubbles in liquids,'' J. Acoust. Soc. Am. **25**, 536–537 (1953).
[58] C. Feuillade and R. W. Nero, ''A viscous-elastic swimbladder model for describing enhanced-frequency resonance scattering from fish,'' J. Acoust. Soc. Am. **103**, 3245–3255 (1998).
[59] N. Levanon, *Radar Principles* (Wiley, New York, 1988).
[60] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994).
[61] W. T. Hogarth and H. T. Johnson, ''Joint Interim Report, Bahamas Marine Mammal Stranding Event of 15–16 March 2000,'' U.S. Department of Commerce, 2001.

Auditory filter shapes for the bottlenose dolphin (*Tursiops truncatus*) and the white whale (*Delphinapterus leucas*) derived with notched noise[a)]

James J. Finneran
Space and Naval Warfare Systems Center, San Diego, Code 235, 53560 Hull Street, San Diego, California 92152-5000

Carolyn E. Schlundt[b)]
Science Applications International Corporation, Maritime Services Division, 3990 Old Town Avenue, Suite 105A, San Diego, California 92110

Donald A. Carder and Sam H. Ridgway
Space and Naval Warfare Systems Center, San Diego, Code 235, 53560 Hull Street, San Diego, California 92152-5000

(Received 7 November 2001; revised 23 March 2002; accepted 30 May 2002)

Auditory filter shapes were estimated in two bottlenose dolphins (*Tursiops truncatus*) and one white whale (*Delphinapterus leucas*) using a behavioral response paradigm and notched noise. Masked thresholds were measured at 20 and 30 kHz. Masking noise was centered at the test tone and had a bandwidth of 1.5 times the tone frequency. Half-notch width to center frequency ratios were 0, 0.125, 0.25, 0.375, and 0.5. Noise spectral density levels were 90 and 105 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$. Filter shapes were approximated using a *roex*(p,r) function; the parameters p and r were found by fitting the integral of the *roex*(p,r) function to the measured threshold data. Mean equivalent rectangular bandwidths (ERBs) calculated from the filter shapes were 11.8 and 17.1% of the center frequency at 20 and 30 kHz, respectively, for the dolphins and 9.1 and 15.3% of the center frequency at 20 and 30 kHz, respectively, for the white whale. Filter shapes were broader at 30 kHz and 105 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$ masking noise. The results are in general agreement with previous estimates of ERBs in *Tursiops* obtained with a behavioral response paradigm. [DOI: 10.1121/1.1488652]

PACS numbers: 43.80.Lb [WA]

I. INTRODUCTION

Studies of marine mammal audition have demonstrated that many of these animals possess keen hearing and sophisticated auditory systems (reviewed by Nachtigall, 1986). Bottlenose dolphins (*Tursiops truncatus*), white whales (*Delphinapterus leucas*), and other echolocating odontocetes have developed especially intricate hearing abilities, including high sensitivity, wide bandwidth, and the ability to detect acoustic signals in noisy environments (Au, 1993). A primary feature of the auditory system in these animals is the ability to resolve a complex sound into its individual frequency components. Frequency selectivity in mammalian auditory systems is often modeled as a collection of overlapping bandpass filters called the ''auditory filters'' (e.g., Fletcher, 1940; Patterson, 1974). These models are based, in part, on observations that acoustic signals are most easily masked by signals at the same or nearby frequencies. The shape and size of the auditory filters affect the loudness and detectability of complex sounds and broadband signals (Scharf, 1970).

The frequency bandwidth of the auditory filters in humans was first investigated by Fletcher (1940), who measured pure-tone hearing thresholds in the presence of variable bandwidth noise. In these measurements, the masked threshold increased with the noise bandwidth until the noise bandwidth exceeded some value, called the ''critical bandwidth.'' At noise bandwidths exceeding the critical bandwidth, the masked thresholds remained constant. From these data Fletcher proposed a simple model for masking consisting of a rectangular filter with a width equal to the critical bandwidth centered at the tone frequency. Fletcher (1940) also proposed a technique for indirectly estimating the critical bandwidth by measuring the masked threshold of a tone in the presence of broadband noise. This technique assumed that at threshold the ratio of tone power to noise power inside the critical bandwidth was constant. For white noise, the noise power within the critical bandwidth is the noise spectral density multiplied by the critical bandwidth; thus, the critical bandwidth could be estimated from the ratio of tone power to noise power spectral density. This ratio is referred to as the critical ratio. Critical ratios in humans are approximately 2.5 times smaller than critical bandwidths (Scharf, 1970). Although critical ratio measurements provide a simple

[a)]The noise generation technique and notched-noise thresholds were presented in ''Technique for the generation and frequency compensation of band-limited white noise and its application in studies of masked hearing thresholds'' and ''Masked hearing thresholds and critical bandwidths for dolphins and a white whale at 20 and 30 kHz,'' respectively, at the 138th Meeting of the Acoustical Society of America (November, 1999).

[b)]Current address: EDO Professional Services, 3270 1/2 Rosecrans St., San Diego, CA 92110.

 0001-4966/2002/112(1)/322/7/$19.00

way to estimate the critical bandwidth (and hence the auditory filter width), the critical ratio is probably more closely related to the processing efficiency of the auditory system, rather than the filter width, and thus is a poor estimator of frequency selectivity (Patterson *et al.*, 1982). Furthermore, the assumption of rectangular filter shapes is not consistent with experimental data showing a smooth and continuous change in threshold as a function of noise bandwidth (e.g., Schafer *et al.*, 1950; Greenwood, 1961; Swets *et al.*, 1962).

Subsequent studies derived human auditory filter shapes from behavioral measures of pure-tone hearing thresholds masked by narrow-band noise (e.g., Patterson, 1974), rippled noise (e.g., Houtgast, 1977), and notched-noise (e.g., Patterson, 1976; Patterson *et al.*, 1982; Glasberg and Moore, 1990). These studies relied upon the power spectrum model of masking, which states that the power of the signal at the masked threshold, P_s, is given by

$$P_s = K \int_{-\infty}^{\infty} N(f) W(f) df, \tag{1}$$

where K is a constant, $N(f)$ is the noise power spectral density, $W(f)$ is a weighting function describing the shape of the auditory filter, and f is the frequency (Glasberg and Moore, 1990). The value of K is a function of the subject, type of masker, and center frequency, and equals the signal-to-noise ratio at the output of the filter required to reach threshold (Glasberg and Moore, 1990). The notched-noise technique has been widely used with human subjects, in part because it limits off-frequency listening (the use of a filter that is not centered at the tone frequency) (Moore *et al.*, 1995). In the notched-noise method, thresholds are measured in the presence of broadband noise containing a spectral notch of varying width. The filter shape is derived by first assuming a mathematical form for W, then fitting the integral of W to the data relating masked threshold to notch width. Often, W is assumed to be described by a rounded exponential (*roex*) function such as

$$W(g) = (1-r)(1+pg)e^{-pg} + r, \tag{2}$$

where g is the frequency deviation from the tone frequency divided by the tone frequency [$g = (f - f_0)/f_0$, where f_0 is the tone frequency] and p and r are adjustable parameters (Patterson *et al.*, 1982). This process yields a continuous function for the auditory filter shape; many authors have used the concept of the equivalent rectangular bandwidth (ERB) to describe the width of the filter using a single numeric value. The ERB is defined as the width of a rectangular filter which has the same peak amplitude and area under the curve as the filter shape (Moore, 1989). The ERB of the filter described by Eq. (2) reduces to $4f_0/p$ for small values of r (Patterson *et al.*, 1982).

Several of the same techniques developed to study human frequency selectivity have been applied to dolphins and white whales. Johnson (1968) and Johnson *et al.* (1989) employed behavioral methods to measure pure-tone thresholds in the presence of broadband noise and used the resulting data to estimate critical ratios in the bottlenose dolphin and white whale, respectively. Au and Moore (1990) used a behavioral response paradigm and variable bandwidth noise to measure critical ratios and critical bandwidths in a bottlenose dolphin at 30, 60, and 120 kHz. Popov *et al.* (1997) used notched-noise and auditory brain-stem response (ABR) recordings to estimate auditory filter shapes in a young bottlenose dolphin at 64, 76, 90, and 108 kHz. Popov *et al.* (1997) derived auditory filter shapes using the *roex*(p,r) filter shape in Eq. (2); the resulting ERBs were approximately 3% of the center frequency over the range 64–108 kHz. These values are much smaller than typical ERBs for humans (Moore and Glasberg, 1983) and the critical bandwidths measured by Au and Moore (1990). Lemonds *et al.* (2000) estimated auditory filter shapes at 40, 60, 80, and 100 kHz in a 20-year-old bottlenose dolphin using notched noise and a behavioral response paradigm. ERBs derived from these data [using the *roex*(p,r) filter shape] ranged from 11%–16% of the center frequency. It is unknown whether the discrepancy between Lemonds *et al.* (2000) and Popov *et al.* (1997) was a result of methodological differences (e.g., ABR compared to behavioral measures), or differences between the subjects themselves.

The current report presents the results of a study to measure auditory filter shapes in the bottlenose dolphin and white whale at 20 and 30 kHz. A behavioral response paradigm was used to measure pure-tone underwater hearing thresholds in the presence of broadband noise containing a spectral notch. Measurements were conducted at two different noise levels and five different notch widths. The resulting data were used to derive auditory filter shapes using the *roex*(p,r) filter shape.

II. METHODS

A. Experimental subjects

Experimental subjects consisted of two bottlenose dolphins (BEN, age 36 years, approximately 250 kg; NEM, age 34, 220 kg) and one white whale (MUK, age 32, 550 kg). The study followed a protocol approved by the Institutional Animal Care and Use Committee at SPAWAR Systems Center, San Diego, and operated under all applicable Department of Defense guidelines.

B. Procedure

The hearing test procedure is described in detail in Schlundt *et al.* (2000) and Finneran *et al.* (2000). Tests were conducted in a 9×9-m floating netted enclosure located in San Diego Bay. The test apparatus consisted of two underwater listening stations, designated as the "S1 station" and the "S2 station" (Fig. 1). The S1 station was the site for the presentation of a "start" signal (1-s duration, 20 kHz, approximately 140 dB $re{:}1\ \mu$Pa) to begin the hearing test. The hearing tests were conducted at the S2 station. Each station consisted of a polyvinylchloride (PVC) frame with a plastic biteplate on which the subjects were trained to position. The S1 and S2 biteplates were located at depths of approximately 4 and 2 m, respectively. The use of two underwater stations was a by-product of previous studies conducted with these test subjects (Schlundt *et al.*, 2000; Finneran *et al.*, 2000),

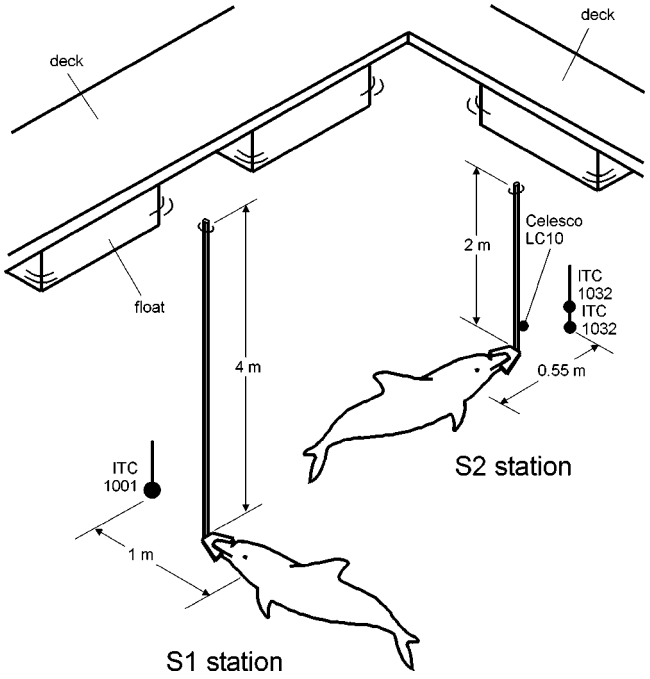


FIG. 1. Experimental setup showing the S1 and S2 stations with a subject in position.

where it was desired to spatially separate the hearing test location (the S2 station) from the site of exposure to a fatiguing sound (the S1 station).

The hearing test procedure was based on the method of free response, or MFR (Egan *et al.*, 1961). Each subject was presented with a number of hearing test tones (i.e., trials) during a relatively long observation period, designated here as a ''dive.'' Each dive began with the trainer directing the subject to the S1 station, where it remained until presented with the S1 start signal. Upon hearing the S1 start signal, the subject proceeded to the S2 station for the hearing test. At the S2 station, hearing test tones were presented to the subject. The time interval between trials was randomized between 5 and 10 s and the subject did not know when the next trial would occur. Subjects were trained to whistle if they heard a tone and to remain quiet otherwise. The time period between 0.05 and 2.05 s immediately following each tone start was designated as a ''hit interval.'' Only whistle responses occurring within a hit interval were recorded as ''hits,'' any whistle response by a subject not occurring within a hit interval was recorded as a false alarm. Thirty percent of the trials were signal absent, or ''catch'' trials. The false-alarm rate R_{FA} was defined as

$$R_{\mathrm{FA}}=\left(\frac{N_{\mathrm{FA}}}{T-N_T T_1}\right)T_1, \tag{3}$$

where N_{FA} is the number of false alarms, T is the total amount of time the subject spent on the S2 station, N_T is the number of tones presented, and T_1 is the hit interval duration. The term in parentheses is the number of false alarms divided by the total amount of time during which the subject was on the S2 station with no hit interval present. This term is multiplied by T_1 to obtain a dimensionless quantity.

The amplitudes of the tones were adjusted using a modified up/down staircase procedure (e.g., Cornsweet, 1962): the amplitude was decreased 2 dB following each hit (a whistle response to a tone) and increased 2 dB following each miss (no whistle response to a tone). After a variable number of trials, the trainer sounded an underwater buzzer which signaled the subject to return to the surface for fish reward. The next dive was then begun, if necessary. Hearing thresholds were calculated as the mean sound-pressure level (SPL) of the first 10 hit–miss/miss–hit reversal points. Thresholds were measured at tone frequencies of 20 and 30 kHz.

The S1 start signals were produced using a programmable function generator (Wavetek 178) and amplified (BGW PS4) before being input to the S1 sound projector (ITC 1001). The hearing test tones were 250 ms in duration including 5-ms rise and fall times. Tones were generated using a personal computer (PC) containing a multifunction board (National Instruments PCI-MIO-16E-1), filtered (Ithaco 4212), and amplified (BGW PS2) before being input to the tone projector (ITC 1032). Received tone levels were calibrated (without the subjects present) using a hydrophone (B&K 8103). The hydrophone output was amplified and filtered (B&K 2635 and Ithaco 4302), then digitized with the PCI-MIO-16E-1. Subject whistle responses were measured and digitized using a hydrophone (Celesco LC-10), amplifier (B&K 2635), and the PCI-MIO-16E-1.

C. Noise stimuli

Masking noise was generated on a dedicated computer running custom software. This system allowed for the continuous generation of masking noise whose frequency spectrum was compensated to eliminate the effects of projector frequency-dependent transmission characteristics and acoustic multipath propagation. Bandlimited white noise with a Gaussian amplitude distribution was first generated using a PCI-MIO-16E-1, attenuated (HP 355D), filtered (SRS 560 and Ithaco 4302), amplified (BGW PS2), then input to the noise projector (ITC 1032). The resulting sound pressure at the location of the subject's ears (without the subject present) was measured using a B&K 8103 hydrophone, amplified and filtered (B&K 2635 and Ithaco 4302), then digitized using an HP 3561A dynamic signal analyzer. The measured pressure spectral density was used to create a digital filter such that, when the initially white noise passed the filter, the resulting pressure spectral density at the subject's location would be zero within the notch and flat within the passbands. This condition required a digital filter with zero amplitude in the notch and a passband amplitude derived from the inverse of the pressure spectral density originally produced by the (unfiltered) white noise. An inverse FFT technique was used to create the digital filter coefficients from the desired filter spectral density. Broadband noise was then continuously generated and applied to this digital filter to produce notched noise with a flat spectral density within the passbands.

Masking noise was centered at the tone frequency f_0 and had a bandwidth of 30 or 45 kHz for tone frequencies of 20 or 30 kHz, respectively. These bandwidths were selected so that the no-notch-noise bandwidths would be larger than the

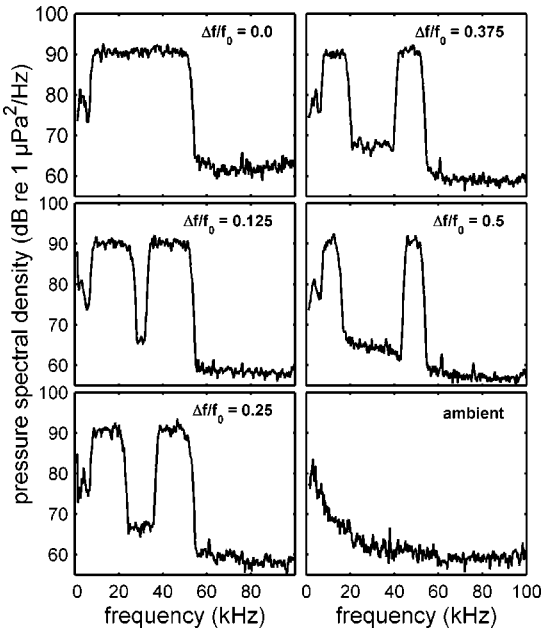


FIG. 2. Representative spectra for 30 kHz, 90 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$ masking noise. Notch widths were $\Delta f/f_0 = 0$, 0.125, 0.25, 0.375, 0.5. The lower right panel shows the ambient noise spectral density measured in San Diego Bay that day. Each plot shows the average of 40 individual spectra computed from 4-ms time series. The frequency resolution was 250 Hz.

critical bandwidth measured by Au and Moore (1990) at 30 kHz, which was 17 kHz.

Noise having different notch widths was described by the ratio of one-half the notch width to the tone frequency ($\Delta f/f_0$). Thresholds were measured using $\Delta f/f_0$ values of 0, 0.125, 0.25, 0.375, and 0.5 for both 20- and 30-kHz tone frequencies. The order at which different notch widths were tested was varied. The pressure spectral density within the passbands was flat within ± 3 dB for these notch widths. Tests were conducted at noise spectral densities of 90 and 105 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$. Thresholds were also measured without masking noise (ambient noise only).

Figure 2 shows some representative notched-noise spectra at 30 kHz center frequency and 90 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$ pressure spectral density level. Each plot is the average of 40 individual spectra computed from 4-ms time series. Figure 2 also shows the ambient noise measured in San Diego Bay. Ambient noise levels were relatively high; thus, within the notch the ambient noise is visible; however, masking noise spectral densities within the passbands were 25–40 dB above the ambient noise spectral density. The ratio of masking noise power to ambient noise power (within the notch) was approximately 22 dB at the largest notch width and lower noise level.

III. RESULTS

Figure 3 shows the masked hearing thresholds (left ordinate) and false-alarm rates (right ordinate) measured for BEN, NEM, and MUK at 20 and 30 kHz as functions of $\Delta f/f_0$. The filled circles and triangles indicate thresholds measured at noise levels of 90 and 105 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$, respectively; the filled squares indicate thresholds measured in the presence of ambient noise only. At zero notch width,

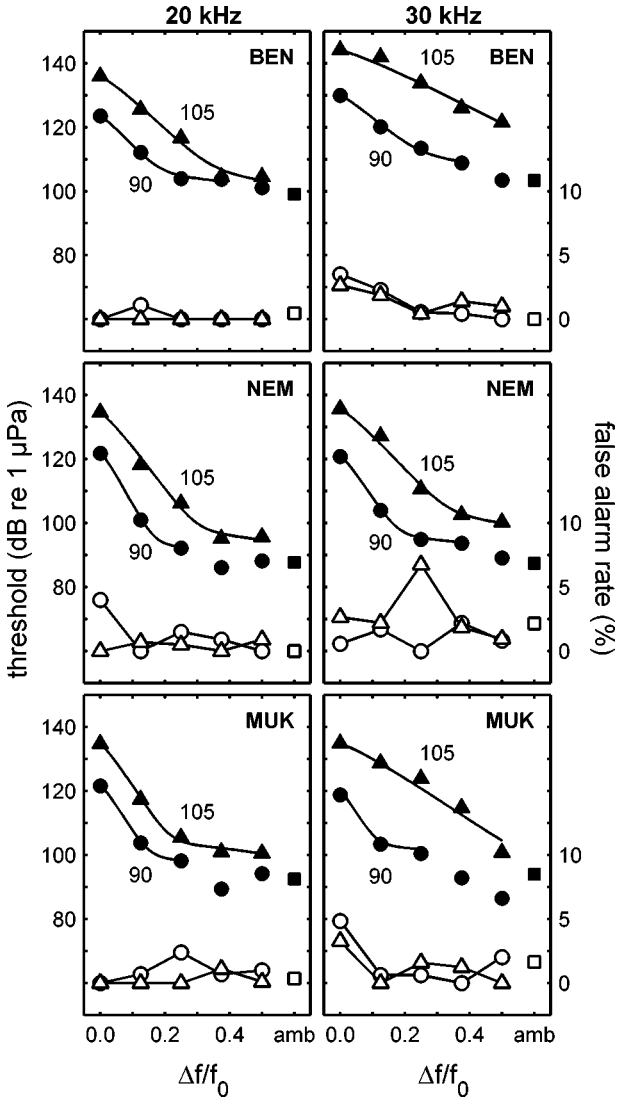


FIG. 3. Masked thresholds (filled symbols) and false-alarm rates (open symbols) measured at 20 kHz (left panels) and 30 kHz (right panels) as functions of $\Delta f/f_0$ for BEN (upper), NEM (middle), and MUK (lower). Circles: 90 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$ masking noise; Triangles: 105 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$ masking noise; Squares: ambient noise only. The solid lines connecting the thresholds were obtained by fitting the integral of the *roex*(p,r) filter function given in Eq. (5) to the measured data. The numbers by each curve indicate the masking noise spectral density level (in dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$).

subjects' thresholds at 90 and 105 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$ were separated by approximately 15 dB as expected (see Hawkins and Stevens, 1950; Yost, 1994). Thresholds decreased as the notch width increased. At large notch widths the data flattened, especially the 20-kHz data, and approached the thresholds measured in ambient noise only. The open circles, triangles, and squares show the false-alarm rates measured at noise levels of 90 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$, 105 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$, and ambient noise only, respectively. False-alarm rates were generally in the range 0%–3%. False-alarm rates at 30 kHz, $\Delta f/f_0 = 0$, were higher for both BEN and MUK.

For the notched-noise used in this study, Eq. (1) simplifies to

$$P_s(\Delta f/f_0) = 2KN_0 f_0 \int_{\Delta f/f_0}^{0.75} W(g)\,dg, \tag{4}$$

TABLE I. *Roex*(p,r) curve-fit parameters at 20 and 30 kHz.

Frequency (kHz)	Subject	90 dB *re*: 1 μPa^2/Hz		105 dB *re*: 1 μPa^2/Hz	
		p	r	p	r
20	BEN	32.3	1.46×10^{-3}	25.5	1.63×10^{-4}
	NEM	50.5	8.51×10^{-5}	35.3	2.26×10^{-5}
	MUK	46.5	3.65×10^{-4}	41.7	6.90×10^{-5}
30	BEN	25.1	1.72×10^{-3}	13.8	4.15×10^{-4}
	NEM	42.5	2.35×10^{-4}	28.6	7.25×10^{-5}
	MUK	49.5	1.22×10^{-3}	17.5	1.40×10^{-5}

where the noise spectral density is assumed to equal N_0 within the passbands and to be zero within the notch. The integration limit of 0.75 is equal to the value of $\Delta f/f_0$ for the upper band of the masking noise. The *roex*(p,r) function [Eq. (2)] was chosen as the mathematical form for $W(g)$; the integral of $W(g)$ is thus $-(1-r)p^{-1}(2+pg)\exp(-pg)+rg$. To compute the filter coefficients p and r, the masked threshold data in Fig. 3 were fitted with the equation

$$L_s(\Delta f/f_0)=K'+10\log_{10}\{[-(1-r)p^{-1}(2+pg)\times\exp(-pg)+rg]|_{\Delta f/f_0}^{0.75}\}, \quad (5)$$

where L_s is the signal SPL at threshold. The constant K' indicates the efficiency of the detection process and is provided by the fitting process; only r and p are adjusted to fit the data. As in Patterson *et al.* (1982) the values of r were constrained to be no less than 20 dB below $W(0.4)$. This condition restricts r to values considered realistically possible and has little effect on the curve fit itself (r values this small have a negligible effect on the curve fit result or the filter shape) (Patterson *et al.*, 1982). Masked thresholds within 3 dB of the thresholds measured in the presence of ambient noise only were excluded from the curve fit. This had no effect on the curve fits to the 105-dB re:1 μPa^2/Hz noise data. The curve-fit results are displayed as solid lines in Fig. 3. The individual values for r and p are listed in Table I.

Figure 4 shows the *roex*(p,r) filter shapes computed from the p and r values in Table I. The filters are assumed to be symmetric; thus, only the upper half of each is shown. The dashed and solid lines indicate filter shapes at 90 and 105 dB re:1 μPa^2/Hz masking noise, respectively. Values of p and r were lower at 105 dB re:1 μPa^2/Hz noise in each condition, resulting in broader filters and ''flattening'' of the tail sections occurring at lower levels (if at all) for the higher noise level. Filters at 30 kHz were also broader than at 20 kHz for each individual subject.

ERBs were calculated from the curve-fit results using ERB$=4f_0/p$ (Patterson *et al.*, 1982). Figure 5 shows the ERB plotted versus the filter center frequency. The circles, squares, and inverted triangles indicate ERBs for BEN, NEM, and MUK, respectively. The open and filled symbols indicate ERBs at 90 and 105 dB re:1 μPa^2/Hz, respectively. Mean ERBs were 11.8% and 17.1% of the center frequency at 20 and 30 kHz, respectively, for the dolphins and 9.1% and 15.3% of the center frequency at 20 and 30 kHz, respectively, for the white whale. The ERBs were generally higher at 30 kHz and were larger at the higher noise level. Also

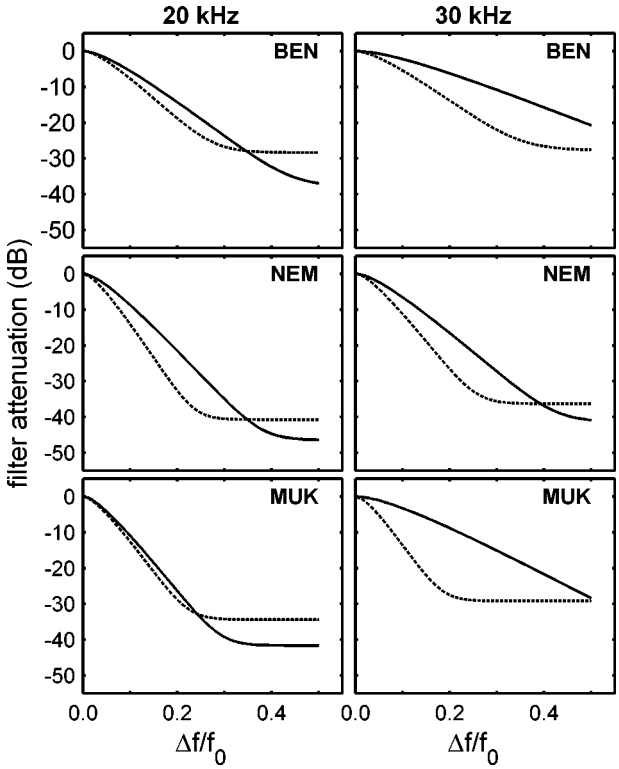


FIG. 4. Auditory filter shapes computed from the curve-fit parameters p and r (Table I). Dashed lines: 90 dB re:1 μPa^2/Hz masking noise; Solid lines: 105 dB re:1 μPa^2/Hz masking noise. The filters are assumed to be symmetric.

shown in Fig. 5 are the critical band data from Au and Moore (1990) (A) and the ERBs reported by Popov *et al.* (1997) (P) and Lemonds *et al.* (2000) (L) for *Tursiops*. Lemonds *et al.* (2000) measured ERBs at 40, 60, 80, and 100 kHz, but only reported ERB values at 40 and 100 kHz. Lemonds *et al.* stated that there was very little change in the ERBs between 60 and 100 kHz; therefore, the ERB at 100 kHz was used for the 60- and 80-kHz data points in Fig. 5. ERBs from the current study at 30 kHz are lower than the critical bandwidths measured by Au and Moore at 30 kHz, but are close to the ERB measured by Lemonds *et al.* at 40 kHz. The current data also appear to follow the trend established with

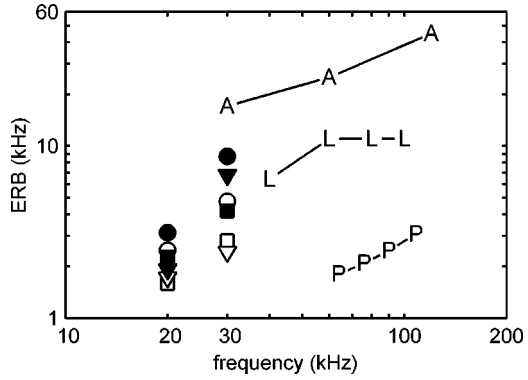


FIG. 5. ERBs and critical bandwidths for dolphins and white whales. Circles: BEN; Squares: NEM; Triangles: MUK; Open symbols: 90 dB re:1 μPa^2/Hz masking noise; Closed symbols: 105 dB re:1 μPa^2/Hz masking noise; A: Critical bandwidths from Au and Moore (1990); L: ERBs from Lemonds *et al.* (2000); P: ERBs from Popov *et al.* (1997).

the Lemonds *et al.* data, that is increasing with frequency below approximately 60 kHz. The ERBs at 30 kHz are generally higher than those measured by Popov *et al.* (1997).

IV. DISCUSSION

The masked thresholds in Fig. 3 follow the relationships predicted by the power spectrum model of masking: as the notch width increased, the noise power within the auditory filter was reduced, lowering the required signal power at threshold. At zero notch width, the threshold curves for each subject/frequency case are roughly parallel and separated by approximately 15 dB, the difference between the noise levels. At larger notch widths, the thresholds approached the thresholds masked by ambient noise only. The filter shape derivation [Eqs. (4) and (5)] assumed that the noise spectral density was constant in the passbands and zero in the notch. Ambient noise levels within the notch were not zero (see Fig. 2); thus, at large notch widths the total noise power at threshold may have been underestimated in the curve-fitting process. This effect would have been more pronounced at the lower masking noise level (90 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$) and largest notch width ($\Delta f/f_0 = 0.5$), where the ratio of masking noise power to ambient noise power (within the notch) was lowest. For this reason, masked thresholds within 3 dB of the ambient noise-only thresholds were excluded from the curve-fit process. This criterion had no impact on the 105 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$ noise data, since these thresholds were well above the ambient-only thresholds. The remaining data were fit well by the *roex*(p,r) filter shape.

Auditory filter shapes in Fig. 4 were broader at 30 kHz compared to those at 20 kHz. Filter shapes at 105-dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$ noise were also broader than those measured with 90 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$ noise. The filters also exhibited a substantial "flattening" of the skirts at large values of $\Delta f/f_0$. This effect was more pronounced at the 90-dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$ noise level and at 20 kHz. Filter skirts at 105-dB noise flattened at larger $\Delta f/f_0$ and lower attenuation values compared to those at 90 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$ noise. This suggests that the flattened skirts are a result of the notched-noise-masked thresholds approaching the thresholds masked by ambient noise only, and are not an inherent part of the auditory filter itself. The relationships between filter shapes measured at different noise levels (i.e., higher noise levels resulted in broader filters, larger dynamic range, less flattening of the skirts) were consistent with auditory filter shapes measured in humans at varying noise levels (Glasberg *et al.*, 1984; Shailer *et al.*, 1990).

The ratio of noise bandwidth (with no notch) to center frequency was 1.5 for the noise centered at 20 and 30 kHz. These bandwidths were selected after a review of the threshold versus noise bandwidth data presented by Au and Moore (1990). The thresholds measured by Au and Moore reached asymptotes at noise bandwidths less than the center frequency, so the use of a bandwidth 1.5 times the center frequency insured a zero-notch bandwidth greater than the critical bandwidth. Some of the wider auditory filters (e.g., BEN at 30 kHz) likely extended beyond the noise band; thus, the use of larger noise bandwidths may have resulted in different measured thresholds. However, since the filter shape estimates were based on fitting Eq. (5) to the measured thresholds only over the actual noise passbands, the larger noise bandwidth should not have significantly changed the filter shape estimates.

Glasberg *et al.* (1984) reported human auditory filter dynamic ranges exceeding 60 dB. The dynamic ranges of the filters in the present study were at least 30–45 dB and were larger for the higher noise level. The smaller dynamic range observed in the current study may have been the result of the subjects' ages [dynamic range in humans tended to decrease with age (Glasberg *et al.*, 1984)] or may have been limited by the level of masking noise used relative to the ambient noise level (a higher noise level or lower ambient noise may have revealed a larger dynamic range).

The filter shapes obtained for the dolphins and white whale were similar in shape and width. Filter shapes for the dolphin BEN were broader than those for the dolphin NEM and whale MUK at every frequency/noise level combination. There was no obvious relationship between filter widths for MUK and NEM—filter widths for MUK were broader at some frequency/noise level combinations and smaller at others. Auditory filters obtained from elderly humans or those with pre-existing hearing loss are typically broader than those measured from younger listeners and/or those with normal hearing (e.g., Patterson *et al.*, 1982; Peters and Moore, 1992; Sommers and Humes, 1993). Unfortunately, the small number of subjects and the subjects' ages (32–36 years) in the current study make it impossible to separate effects of intersubject variability from possible changes in filter width caused by aging or pre-existing hearing loss.

ERBs calculated from the current data increase with frequency and agree with those reported for *Tursiops* by Lemonds *et al.* (2000). These ERBs are substantially larger than those reported by Popov *et al.* (1997). The difference in ERBs may be due to the methodology employed or the subject ages. Both the current study and Lemonds *et al.* used a behavioral response paradigm to measure hearing thresholds, while Popov *et al.* used ABR measurements. ABR test tones are typically short-duration tone pips which may have a broad frequency spectrum (tone pips used by Popov *et al.* had ERBs of 1.5 kHz) and potentially confound measurements of frequency selectivity. Also, Popov *et al.* employed a relatively small water-filled box, with the dorsal part of the subject's body above the water. Although the digitally generated noise used by Popov *et al.* possessed a flat frequency spectrum within the passbands, the actual received noise spectrum (at the subject's location) is unknown; it is possible that the enclosed sound field resulted in complex acoustic conditions that produced nonflat noise within the passbands.

V. CONCLUSIONS

(1) Masked threshold data as a function of $\Delta f/f_0$ were fit well by the integral of the *roex*(p,r) filter shape.
(2) Filter shapes at 105 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$ noise were broader than those at 90 dB $re{:}1\ \mu\mathrm{Pa}^2/\mathrm{Hz}$ noise.
(3) Filter shapes at 30 kHz were broader than those measured at 20 kHz.
(4) ERBs were 11.8% and 17.1% of the center frequency at

20 and 30 kHz, respectively, for the dolphins, and 9.1% and 15.3% of the center frequency at 20 and 30 kHz, respectively, for the white whale.

ACKNOWLEDGMENTS

Tricia Kamolnick, Randall Dear, Jennifer Carr, Mark Todd, Linda Green, Scott Peluso, and Debbie Skaar trained the test subjects. We thank Dr. Robert Gisiner and Dr. Frank Stone for support and encouragement. Support was provided by the Office of Naval Research and CNO (N45). James Finneran was supported by a National Research Council Research Associateship.

Au, W. W. L., and Moore, P. W. B. (**1990**). ''Critical ratio and critical bandwidth for the Atlantic bottlenose dolphin,'' J. Acoust. Soc. Am. **88**, 1635–1638.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer, New York).

Cornsweet, T. N. (**1962**). ''The staircase method in psychophysics,'' Am. J. Psychol. **75**, 485–491.

Egan, J. P., Greenberg, G. Z., and Schulman, A. I. (**1961**). ''Operating characteristics, signal detectability, and the method of free response,'' J. Acoust. Soc. Am. **33**, 993–1007.

Finneran, J. J., Schlundt, C. E., Carder, D. A., Clark, J. A., Young, J. A., Gaspin, J. B., and Ridgway, S. H. (**2000**). ''Auditory and behavioral responses of bottlenose dolphins (*Tursiops truncatus*) and white whales (*Delphinapterus leucas*) to impulsive sounds resembling distant signatures of underwater explosions,'' J. Acoust. Soc. Am. **108**, 417–431.

Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–61.

Glasberg, B. R., Moore, B. C. J., Patterson, R. D., and Nimmo-Smith, I. (**1984**). ''Dynamic range and asymmetry of the auditory filter,'' J. Acoust. Soc. Am. **76**, 419–427.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

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

Hawkins, J. E., and Stevens, S. S. (**1950**). ''The masking of pure tones and of speech by white noise,'' J. Acoust. Soc. Am. **22**, 6–13.

Houtgast, T. (**1977**). ''Auditory filter characteristics derived from direct-masking data and pulsation-threshold data with a rippled-noise masker,'' J. Acoust. Soc. Am. **62**, 409–415.

Johnson, C. S. (**1968**). ''Masked tonal hearing thresholds in the bottlenosed porpoise,'' J. Acoust. Soc. Am. **44**, 965–967.

Johnson, C. S., McManus, M. W., and Skaar, D. (**1989**). ''Masked tonal hearing thresholds in the beluga whale,'' J. Acoust. Soc. Am. **85**, 2651–2654.

Lemonds, D. W., Au, W. W. L., Nachtigall, P. E., and Roitblat, H. L. (**2000**). ''High-frequency auditory filter shapes in an Atlantic bottlenose dolphin,'' J. Acoust. Soc. Am. **108**, 2614(A).

Moore, B. C. J. (**1989**). *An Introduction to the Psychology of Hearing* (Academic, London).

Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,'' J. Acoust. Soc. Am. **74**, 750–753.

Moore, B. C. J., Glasberg, B. R., van der Heijden, M., Houtsma, A. J. M., and Kohlrausch, A. (**1995**). ''Comparison of auditory filter shapes obtained with notched-noise and noise-tone maskers,'' J. Acoust. Soc. Am. **97**, 1175–1182.

Nachtigall, P. E. (**1986**). ''Vision, audition, and chemoreception in dolphins and other marine mammals,'' in *Dolphin Cognition and Behavior: A Comparative Approach*, edited by R. J. Schusterman, J. A. Thomas, and F. G. Wood (Erlbaum, London), pp. 79–113.

Patterson, R. D. (**1974**). ''Auditory filter shape,'' J. Acoust. Soc. Am. **55**, 802–809.

Patterson, R. D. (**1976**). ''Auditory filter shapes derived with noise stimuli,'' J. Acoust. Soc. Am. **59**, 640–655.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). ''The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold,'' J. Acoust. Soc. Am. **72**, 1788–1803.

Peters, R. W., and Moore, B. C. J. (**1992**). ''Auditory filter shapes at low center frequencies in young and elderly hearing-impaired subjects,'' J. Acoust. Soc. Am. **91**, 256–266.

Popov, V. V., Supin, A. Y., and Klishin, V. O. (**1997**). ''Frequency tuning of the dolphin's hearing as revealed by auditory brain-stem response with notch-noise masking,'' J. Acoust. Soc. Am. **102**, 3795–3801.

Schafer, T. H., Gales, R. S., Shewmaker, C. A., and Thompson, P. O. (**1950**). ''The frequency selectivity of the ear as determined by masking experiments,'' J. Acoust. Soc. Am. **22**, 490–496.

Scharf, B. (**1970**). ''Critical bands,'' in *Foundations of Modern Auditory Theory*, edited by J. Tobias (Academic, New York), Vol. I.

Schlundt, C. E., Finneran, J. J., Carder, D. A., and Ridgway, S. H. (**2000**). ''Temporary shift in masked hearing thresholds (MTTS) of bottlenose dolphins, *Tursiops truncatus*, and white whales, *Delphinapterus leucas*, after exposure to intense tones,'' J. Acoust. Soc. Am. **107**, 3496–3508.

Shailer, M. J., Moore, B. C. J., Glasberg, B. R., Watson, N., and Harris, S. (**1990**). ''Auditory filter shapes at 8 and 10 kHz,'' J. Acoust. Soc. Am. **88**, 141–148.

Sommers, M. S., and Humes, L. E. (**1993**). ''Auditory filter shapes in normal-hearing, noise-masked normal, and elderly listeners,'' J. Acoust. Soc. Am. **93**, 2903–2914.

Swets, J. A., Green, D. M., and Tanner, Jr., W. P. (**1962**). ''On the width of critical bands,'' J. Acoust. Soc. Am. **34**, 108–113.

Yost, W. A. (**1994**). *Fundamentals of Hearing*, 3rd ed. (Academic, San Diego).

Changes in auditory sensitivity with depth in a free-diving California sea lion (*Zalophus californianus*)

David Kastak and Ronald J. Schusterman
Long Marine Laboratory, University of California, 100 Shaffer Road, Santa Cruz, California 95060

(Received 19 November 2001; revised 23 April 2002; accepted 3 May 2002)

All current data on underwater hearing in pinnipeds are based on tests conducted in small tanks, and may not accurately represent the auditory functioning of free-ranging animals, especially if hearing sensitivity changes with water depth. Underwater auditory thresholds were determined for a California sea lion at depths ranging from 10 to 100 meters. The following results were obtained: (1) False alarm probabilities (responding in the absence of a signal) decreased significantly with depth, indicating that the sea lion adopted a more conservative response criterion in deeper water. (2) Hearing sensitivity generally worsened with depth. (3) There was a significant interaction between depth and frequency, the depth effect being most pronounced at 10 kHz and reversing at 35 kHz. Increasing pressure related to diving probably alters the impedance characteristics of the pinniped ear, in particular affecting the size of the middle-ear air space via expansion of cavernous tissue in the middle-ear cavity. These results show that the middle ear plays a functional role in underwater sound detection in sea lions. However, contrary to previous speculation, the presence of cavernous tissue in the sea lion middle ear does not appear to enhance sensitivity at depth. [DOI: 10.1121/1.1489438]

PACS numbers: 43.80.Lb [WA]

I. INTRODUCTION

One of the distinctive features of the middle ear of seals and sea lions is the presence of a thickened mucosa with three distinct layers, thought to engorge with blood when submerged. The cavernous tissue of the pinniped middle ear has been thoroughly described by Tandler (1899; harbor seal), Odend'hal and Poulter (1996; California sea lion), Møhl (1968; harbor seal), and Welsch and Riedelsheimer (1997; Weddell seal). The function of this middle-ear mucosal layer has been suggested by these authors as (1) regulating middle-ear volume at depth by expanding under increased static pressure, through either passive or active means, and (2) allowing for movement of the middle-ear ossicles at depth, by ensuring that they were maintained in an air-filled space upon submergence (i.e., the epitympanic recess would be the last reservoir of air in the middle ear as the cavernous tissue expanded during a dive). Møhl (1968) first hypothesized that the cavernous tissue played a potential role in acoustic impedance switching, allowing the pinniped ear to function amphibiously. In air, for example, the input impedance of the middle ear would approximate that of air; in water, the inflation of cavernous tissue would alter the middle-ear impedance to equal that of the aquatic environment. Although the sound conduction mechanisms between the environment and the pinniped inner ear have not been elucidated in water or in air, Repenning (1972) believed that expansion of the cavernous tissue on both sides of the tympanic membrane would allow sound transmission along a pathway identical to that of terrestrial mammals (i.e., tympanic–ossicular transmission), based on a similar concept of impedance matching due to cavernous tissue expansion.

The role of the middle ear in underwater hearing is unknown for terrestrial, marine, or amphibious mammals. In air, the middle ear is commonly thought to serve as an impedance-matching device, converting airborne sound waves (low-pressure, high-particle velocity) into fluid waves (high-pressure, low-particle velocity) in the cochlea. In water, the air-filled middle ear is generally described as an impediment to the transfer of sound energy, because of the air–water barrier at the tympanic membrane, where theoretically, almost all incident sound energy is reflected rather than transmitted (Repenning, 1972). As a result, mammals are thought to hear primarily through bone conduction when under water, ''bone conduction'' referring to any and all sound conduction pathways not utilized in the normal (i.e., outer to middle to inner ear) sense. Vibration of the head, transmission of pressure or flexural waves through the bones of the skull, conduction of particle motion from the environment through the skull to the inner ear, and compression of the cochlear capsule have all been proposed as bone conduction mechanisms (Tonndorf, 1972). The retention of air spaces in the external meatus and middle ear has been viewed solely as a barrier to sound conduction and thus, an impediment to underwater hearing (but see Lipatov, 1992 for a discussion of the role of intrameatal air on hearing sensitivity in pinnipeds).

In contrast to humans, pinnipeds are generally more sensitive to underwater sound than to airborne sound across the audible frequency range (Hamilton, 1957; Hollien and Feinstein, 1975; Schusterman, 1981—for a review of underwater hearing in humans, see Kirkland *et al.*, 1989). Additionally, when tested in water, pinnipeds also have shown an extended upper-frequency hearing limit compared to the upper limit shown in air. One reason for the increase in upper-frequency limit may be related to changes in acoustic impedance—increasing the stiffness of the ossicular suspension upon submergence ought to facilitate high-frequency sound transmis-

sion through the middle ear. As Repenning (1972) notes, there is no evidence that the pinniped middle-ear bones do not function in a relatively fluid environment (i.e., partially or wholly surrounded by cavernous tissue); therefore, if cavernous tissue surrounds the ossicles and contacts both sides of the tympanic membrane at depth, then acoustic energy should be transmitted across the tympanic membrane with the same efficiency as when transmitted from air to the ear. Thus, hearing sensitivity would be better than that expected for a terrestrial ear (Repenning, 1972). These changes in underwater sensitivity should directly reflect the degree to which the cavernous tissue is inflated, and therefore, hearing sensitivity in pinnipeds should change with depth.

To date, all pinniped auditory thresholds have been obtained in shallow tanks, as has all work on masking and the effects of noise. However, conclusions drawn from work conducted in shallow tanks are suspect because of the presence of reflective barriers, pressure release surfaces, and near-field effects. Free-field relationships between sound pressure and particle velocity do not apply in such confined situations, so estimates of acoustic impedance and sound intensity are unlikely to be close approximations of the experimental conditions. Testing under conditions more closely resembling a free field (i.e., open ocean) would provide a means to ecologically validate results obtained in the laboratory. Another advantage of open ocean work relates to the effects of noise on marine mammals: In order to make rational policy decisions regarding anthropogenic noise limits, it is necessary to determine how realistic our estimates of noise exposure for these animals are, given that the work in shallow tanks might over- or underestimate sensitivity of free-ranging (diving) animals to noise.

To resolve the role of the middle ear in hearing at depth, and to provide a comparison open-water study of hearing in pinnipeds, we trained a free-diving male California sea lion to report detection of acoustic signals at depths of 10, 50, and 100 m. We obtained auditory thresholds at each of these depths over a range of frequencies to determine whether pressure-related changes in auditory sensitivity occur.

II. METHODS

A. Subject

The subject of this experiment was ''Newman'' (ZC701), a 12-year-old male California sea lion (*Zalophus californianus*), housed in an open ocean pen at Space and Naval Warfare Systems Center San Diego in San Diego Bay, California. He was fed a daily diet of freshly thawed herring (*Clupea spp.*), capelin (*Mallotus villosus*), smelt, (*Osmarus spp.*) and squid (*Ilex spp.*) totaling about 4 to 5 kg. He consumed about 25%–50% of his daily ration during experimental sessions.

B. Apparatus

The response apparatus consisted of a platform comprising a 31-cm-diameter PVC tube on which were mounted a bite plate, response paddle, and video camera, as shown in Fig. 1. The bite plate provided a stationing area for the subject, who could maintain an invariant head position at a particular depth in the water column by biting on to the neoprene-covered bite-plate surface. The response paddle was an aluminum target plate positioned to the left side of the bite plate. The video camera provided a view of the bite plate and response paddle from above. A dive light, used to delineate trial intervals, was positioned in front of the bite plate and in view of the subject while stationed. A projecting hydrophone (ITC 1032) was suspended in the center of the tube, and an identical receiving hydrophone was mounted on the bite plate. The receiving hydrophone was positioned for calibration purposes in the space that would be occupied by the subject's head during test sessions, and could be pivoted away for testing. The projecting hydrophone was positioned in front of and above the subject, at a distance of 80 cm from the center of the subject's head.

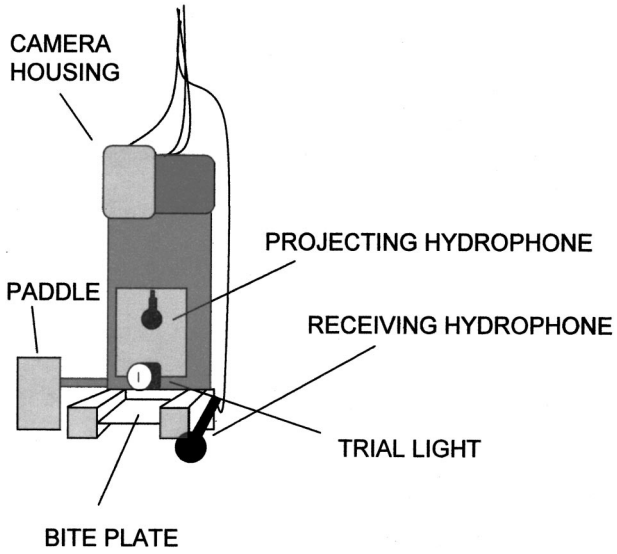


FIG. 1. Diagram of the apparatus for obtaining thresholds at depth from a California sea lion.

The apparatus was suspended via cable from the side of a 26-ft. research vessel, modified to house the equipment for signal production and audio and video monitoring. Test signals (500-ms duration, sampled at 500 kHz, 5-ms rise/fall time, manually triggered) were generated by an experimenter on a personal computer, attenuated, amplified, and projected from the transmitting hydrophone. The incoming signal and ambient noise from the receiving hydrophone were bandpass filtered, amplified, and analyzed using an oscilloscope and a PC-based real-time spectrum analyzer. The dive light was operated manually from the control booth. The trainer monitored live video feed on a constant basis during training and testing. The procedure was blind; that is, the experimenter could not see when a response was made, and the trainer did not know when a signal was triggered. The subject's responses were relayed by the trainer to the experimenter. The experimenter subsequently informed the trainer whether the response was correct or incorrect.

C. Training

The sea lion was trained via classical and instrumental conditioning procedures to dive to the bite plate in shallow water and hold on for a period of time ranging from several

seconds to just over a minute. Subsequently, the subject was trained to respond to a 2-s duration 5-kHz pure tone by pressing the paddle with its nose. In the training phase, we used a discrete trial procedure; after every presentation of a tone, the subject was recalled to the surface using a pinger that produced a series of brief 10-kHz pulses. After acquisition of the stationing and paddle press response, a dive light was introduced to delineate 4-to 6-s trial intervals. At this point in training, a signal could only occur during trial intervals, when the light was turned on. To control for false reporting during the psychophysical phase of the experiment, catch trials, during which the dive light was switched on for 4 to 6 s, but no signal was delivered, were introduced at this time. The behavior of maintaining position (rather than pressing the paddle) was reinforced on catch trials. Following each correct response, the subject was provided feedback by being presented an acoustic conditioned reinforcer (a buzz delivered under water from a small, sealed PVC enclosure).

Following acquisition of single-trial responses, multiple trials were grouped together within a single dive, the subject being trained to restation rather than to surface between trials. Training continued in shallow water until the subject reliably performed ten trials without surfacing until recalled. Fish reinforcement was delivered following completion of a pseudorandomly predetermined number of trials. The amount of reinforcement was determined by the overall amount of correct responding during a trial sequence.

Subsequent to completion of the auditory detection task, the sea lion was trained to enter a cage for transport via the research vessel to the experimental site in 250 m of water, and approximately 10 km off the San Diego coast. Acclimation to open ocean testing took place in stages starting in shallow water near the subject's home pen. Gradually, training moved into San Diego Bay up to depths of about 10 m. The training phase ended following demonstration of normal testing behavior (a series of four dives without spontaneous surfacing) at a depth of 100 m.

D. Psychophysics

Audiometric data were obtained using the go/no-go procedure described above, and a staircase psychophysical method (Stebbins, 1970). The sequence of signal and catch trials was predetermined. For each session, the initial signal level was well above threshold. This signal was attenuated by 4 dB following each HIT (correct detection) until the first MISS (failure to detect), which defined the first reversal point. Subsequently, the signal level was increased by 2 dB following each MISS and decreased by 2 dB following each HIT. Each change of direction in a sequence of signal amplitudes defined a reversal, and sessions comprised a minimum of seven and a maximum of ten reversals. False alarm rates were calculated as the number of FALSE ALARMS (responding in the absence of a signal) divided by the total number of catch trials. Thresholds defined as signal levels corresponding to 50% correct detections were calculated as the mean of the reversal points (Dixon and Mood, 1948).

The thresholds determined by the staircase procedure were ultimately transformed to constant d' thresholds. This calculation adjusted the initial threshold determination based on the number of false alarms per session. Thus, all sensitivity data reported herein are thresholds corresponding to a d' value of 1.

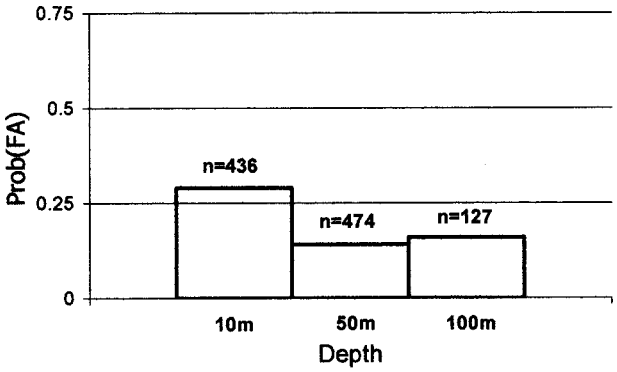


FIG. 2. Proportion of false alarm responding $P(\mathrm{FA})$ vs depth. $P(\mathrm{FA})$ was calculated by pooling training phase data and dividing the number of false alarms by the number of catch trials for a given depth. False alarm responding was significantly greater at 10-m depth than at 50 and 100 m.

Auditory sensitivity at 10 and 50 m was measured at frequencies of 2.5, 6, 10, and 35 kHz. Additionally, thresholds at 100 m were obtained at frequencies of 2.5 and 6 kHz. However, because of small numbers of reversals and high variability, the threshold data at 100 m were excluded from comparison with shallow-water thresholds. The testing order was randomized with respect to depth and frequency in order to eliminate potentially confounding practice effects. The only constraints on testing at particular depths and frequencies on a given day were ambient noise levels and weather conditions. If noise levels were high enough to mask the test signal, dive sequences were run, but the data were not included in threshold calculations. A minimum of two and a maximum of six sessions were used to calculate thresholds at each depth/frequency combination.

III. RESULTS

During training and testing, the subject's response bias changed with depth. Results of pooled training sessions at depths of 10, 50, and 100 m are shown in Fig. 2. False alarm rates were double for sessions conducted at 10-m depth relative to those conducted at 50-and 100-m depth. There were no significant differences in false alarm responding between 50 and 100 m.

Figure 3 shows auditory thresholds (± 1 standard deviation) obtained at depths of 10 and 50 m. Included in this figure are data from Schusterman *et al.* (1972) obtained from a 5–6-year-old male California sea lion in a shallow tank at a depth of a 1 m. These data fairly closely match the 10-m data obtained in the present study. As expected, thresholds varied significantly with frequency, resulting in the typical ''u-shaped'' audiogram. There was also a significant effect of depth on thresholds ($F_{1,26}=12.25$; $p<0.01$) and a significant interaction between depth and frequency ($F_{3,26}=11.05$; $p<0.01$); thresholds were lower in shallow water, except at the highest frequency tested (35 kHz), where this trend was reversed.

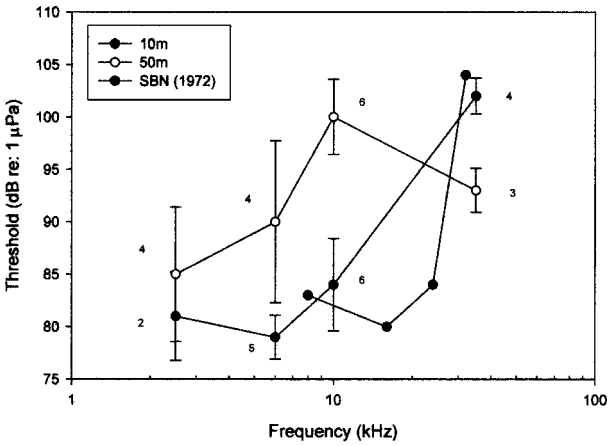


FIG. 3. California sea lion audiograms at 10 and 50 m. Error bars represent standard deviations. Numbers alongside data points represent numbers of sessions conducted at each frequency-depth combination. The plotted points without error bars are reported values for a California sea lion in a small pool (SBN=Schusterman *et al.*, 1972).

IV. DISCUSSION

The California sea lion tested in this study showed a clear tendency to withhold responding at depth, the reasons for which are unclear. It is possible that the subject was more comfortable near the surface, while within view of the boat, and was simply more likely to release the bite plate in order to respond when it was closer to the surface. At depth, if the sea lion were fearful or uncomfortable, it may have been less likely to let go of the bite plate in general, thus less likely to respond in the presence of a signal. The subject's more conservative behavior at depth may also have been a response to the generally increasing task difficulty associated with deeper diving and longer transit time to and from the apparatus at depths of 50 and 100 m. Whatever the reason for the subject's conservative behavior, the difference in response bias among the three depths tested would have had a small but significant effect on estimates of sensitivity had the traditional measure of the 50%-correct detection level been used to define a threshold. Nearly all threshold estimates at 50-m depth would have been high, while estimates obtained at 10 m would have been low, exaggerating the differences evident in the data shown in Fig. 3. It is also possible that small sample sizes for estimating false alarm rates (20 to 30 catch trials per session) led to poor estimates of detectability. However, such errors would have had to be extremely large in magnitude to account for differences of 10 dB or more. Because the 10-m thresholds obtained in this study were similar to those obtained by Schusterman *et al.* (1972) for the same species in a shallow tank, we conclude that audiometric data collected from captive subjects are probably fairly representative of hearing capabilities of free-ranging animals. However, based on the results at depth, this comparison may be valid *only* for shallow water.

Physiological mechanisms must be taken into consideration as the primary cause of the sensitivity changes with depth. It may make little sense to think of the air–water interface at the tympanic membrane as a reflective barrier to the transmission of sound, especially at low frequencies, where the sound wavelengths are many times larger than the structures of the ear. If, at these frequencies, the middle-ear structures act as a damped resonant air space, transmitting acoustic particle motion directly to the inner ear, then changes in pressure, density, and volume would be expected to have frequency-dependent effects on hearing sensitivity. The interface between the environment and the middle-ear space might only become important as a barrier to sound transmission at high frequencies where the wavelengths are similar in size to the middle ear, external meatus, and other tissue and bony structures surrounding the ear. Thus, in shallow water, high-frequency sensitivity would be relatively poor, because of the significant reflection of sound energy at the middle ear. In deeper water, sensitivity might improve because of better impedance matching due to the expansion of cavernous tissue on both sides of the tympanic membrane.

Two separate types of change in the middle ear might account for the patterns of sensitivity change that occurred with depth. At low frequencies, changes in the resonance properties of the middle-ear space should affect sound transmission in a frequency-dependent manner. At high frequencies, expansion of cavernous tissue increases the efficiency of sound transmission through the ossicular pathway, also leading to frequency-dependent alterations in sound transmission with depth, but in the opposite direction. These proposed mechanisms doubtless oversimplify the problem of underwater hearing in pinnipeds, and do not take into consideration sound channels through tissue and bone or the possible role of the middle-ear ossicles acting as a load on the inner ear (which might be stimulated through some other mechanism altogether). However, these ideas are consistent with the results of this study and imply not only that high-frequency sensitivity improves with depth in the California sea lion, but that the upper-frequency hearing limit might also increase with depth. It is interesting to note that in a recent study by Ridgway *et al.* (2001), two belugas tested at depths of up to 300 m showed no evidence of changes in hearing sensitivity at frequencies ranging from 500 Hz to 100 kHz. These authors concluded that in the beluga, underwater hearing does not require the usual ossicular sound conduction/amplification found in terrestrial mammals. The odontocete middle ear, however, is highly specialized compared to that of the pinniped, and differences in structural anatomy and sound conduction pathways are likely to account for the different effects of depth on hearing in these two taxa.

It is evident from this study that the sea lion middle ear is functional under water, in depth- and frequency-dependent ways. Future investigations of the effects of pressure on hearing should focus on several problems. First, experimental variables such as temporal patterns of pressure change, trial and session length, and response bias should be better controlled. These problems might be addressed by testing in a hyperbaric chamber, where the time taken to simulate descent to various depths could be controlled. Second, the physiological and anatomical correlates of diving should be monitored. This might take place through ultrasound or magnetic resonance imaging when a dive response can be induced in sedated or trained subjects, answering the question of whether (a) middle-ear cavernous tissue does engorge, and

(b) whether blood flow to the middle ear is actively regulated (as part of the dive response) or passively controlled, in response to pressure changes. Finally, similar experiments should be conducted with the true seals, or phocids, which generally have larger middle-ear spaces, more extensive cavernous tissue, and are deeper divers. Only after data addressing these issues are obtained will the mechanisms of underwater auditory functioning in pinnipeds be fully understood.

ACKNOWLEDGMENTS

The work reported herein was funded by a grant from the Office of Naval Research to R.J.S. The authors would like to thank Sam Ridgway, Don Carder, Randy Brill, and Pat Moore for technical and logistical support. This research was conducted with the support and cooperation of the United States Navy Marine Mammal Program and Space and Naval Warfare Systems Center, San Diego, and was conducted under protocols approved by the Institutional Animal Care and Use Committees of SPAWAR and UCSC. Steven Meck provided training and boat-handling expertise, and contributed valuable input to experimental design and analysis. We thank Colleen Reichmuth Kastak, Brandon Southall, and two anonymous reviewers for providing critically helpful suggestions on earlier versions of this manuscript.

Dixon, W. J., and Mood, A. M. (**1948**). ''A method for obtaining and analyzing sensitivity data,'' J. Am. Stat. Assoc. **43**, 109–126.

Hamilton, P. M. (**1957**). ''Underwater hearing thresholds,'' J. Acoust. Soc. Am. **29**, 792–794.

Hollien, H., and Feinstein, S. (**1975**). ''Contribution of the external auditory meatus to auditory sensitivity underwater,'' J. Acoust. Soc. Am. **57**, 1488–1492.

Kirkland, P. C., Pence, E. A., Dobie, R. A., and Yantis, P. A. (**1989**). ''Underwater noise and the conservation of divers' hearing: A review,'' Technical report APL-UW TR8930, University of Washington.

Lipatov, N. V. (**1992**). ''Underwater hearing in seals: The role of the outer ear,'' in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Ya. Supin (Plenum, New York), pp. 249–256.

Møhl, B. (**1968**). ''Hearing in seals,'' in *The Behavior and Physiology of Pinnipeds*, edited by R. J. Harrison, R. C. Hubbard, R. S. Peterson, C. E. Rice, and R. J. Schusterman (Appleton-Century-Crofts, New York), pp. 172–195.

Odend'hal, S., and Poulter, T. C. (**1966**). ''Pressure regulation in the middle ear cavity of sea lions: A possible mechanism,'' Science **153**, 768–769.

Repenning, C. A. (**1972**). ''Underwater hearing in seals: Functional morphology,'' in *Functional Anatomy of Marine Mammals*, edited by R. J. Harrison (Academic, London), Vol. 1, pp. 307–331.

Ridgway, S. H., Carder, D. A., Kamolnick, T., Smith, R., Schlundt, C. E., and Elsberry, W. R. (**2001**). ''Hearing and whistling in the deep sea: Depth influences whistle spectra but does not attenuate hearing by white whales (*Delphinapterus leucas*) (Odontoceti Cetacea),'' J. Exp. Biol. **204**, 3829–3841.

Schusterman, R. J. (**1981**). ''Behavioral capabilities of seals and sea lions: A review of their hearing, visual, learning and diving skills,'' Psychol. Record **31**, 125–143.

Schusterman, R. J., Balliet, R. F., and Nixon, J. (**1972**). ''Underwater audiogram of the California sea lion by the conditioned vocalization technique,'' J. Exp. Anal Behav. **17**, 339–350.

Stebbins, W. C. (**1970**). ''Principles of animal psychophysics,'' in *Animal Psychophysics: The Design and Conduct of Sensory Experiments*, edited by W. C. Stebbins (Appleton-Century-Crofts, New York), pp. 1–19.

Tandler, J. (**1899**). ''Ueber ein corpus cavernosum tympanicum beim seehund,'' Monatsschr. Ohrenheilk. **33**, 437–440.

Tonndorf, J. (**1972**). ''Bone conduction,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), pp. 195–237.

Welsch, U., and Riedelsheimer, B. (**1997**). ''Histophysiological observations on the external auditory meatus, middle, and inner ear of the Weddell seal (*Leptonychotes weddelli*),'' J. Morphol. **234**, 25–36.

Audiogram of a harbor porpoise (*Phocoena phocoena*) measured with narrow-band frequency-modulated signals

Ronald A. Kastelein,[a] Paulien Bunskoek, and Monique Hagedoorn
Harderwijk Marine Mammal Park, Strandboulevard-oost 1, 3841 AB Harderwijk, The Netherlands[b]

Whitlow W. L. Au
Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734

Dick de Haan
Netherlands Institute for Fisheries Research (RIVO-DLO), P.O. Box 68, 1970 AB IJmuiden, The Netherlands

(Received 12 June 2001; accepted for publication 27 March 2002)

The underwater hearing sensitivity of a two-year-old harbor porpoise was measured in a pool using standard psycho-acoustic techniques. The go/no-go response paradigm and up–down staircase psychometric method were used. Auditory sensitivity was measured by using narrow-band frequency-modulated signals having center frequencies between 250 Hz and 180 kHz. The resulting audiogram was U-shaped with the range of best hearing (defined as 10 dB within maximum sensitivity) from 16 to 140 kHz, with a reduced sensitivity around 64 kHz. Maximum sensitivity (about 33 dB *re* 1 μPa) occurred between 100 and 140 kHz. This maximum sensitivity range corresponds with the peak frequency of echolocation pulses produced by harbor porpoises (120–130 kHz). Sensitivity falls about 10 dB per octave below 16 kHz and falls off sharply above 140 kHz (260 dB per octave). Compared to a previous audiogram of this species (Andersen, 1970), the present audiogram shows less sensitive hearing between 2 and 8 kHz and more sensitive hearing between 16 and 180 kHz. This harbor porpoise has the highest upper-frequency limit of all odontocetes investigated. The time it took for the porpoise to move its head 22 cm after the signal onset (movement time) was also measured. It increased from about 1 s at 10 dB above threshold, to about 1.5 s at threshold. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1480835]

PACS numbers: 43.80.Lb, 43.80.Ev, 43.80.Jz [FD]

I. INTRODUCTION

The harbor porpoise (*Phocoena phocoena*) is one of the smallest cetacean species and has a relatively wide distribution (Gaskin, 1992). All over its distribution area it is accidentally caught in fisheries (Northridge, 1991). Attempts are being made to reduce this bycatch of harbor porpoises by deterring them with underwater acoustic alarms (pingers). In some cases this technique seems promising (Lien *et al.*, 1995; Kraus *et al.*, 1997; Laake *et al.*, 1998; Trippel *et al.*, 1999; Gearin *et al.*, 2000; Culik *et al.*, 2000). However, much research is still needed to refine this technique, and to determine if habituation occurs, and if so, how it can be avoided (Kastelein *et al.*, 1997, 2000, 2001). Some current questions are: how many alarms are needed to deter harbor porpoises from a net? Is the best approach one alarm with a high source level (SL) in the center of the net, or several alarms with lower SLs distributed over the length of the net? What is the minimal distance at which a porpoise should be deterred, and what should the SLs be under various ambient noise levels? Parameters to consider in relation to SL and number of alarms are: acoustic pollution of the oceans, costs, energy demand of the pingers, efficiency in reducing bycatch, availability, and ease of use.

The first step to answer these questions is to acquire fundamental information about the underwater hearing sensitivity of harbor porpoises. The underwater hearing of the harbor porpoise has been previously studied behaviorally by Andersen (1970) and by means of electrophysiological tests (Bibikov, 1992; Popov *et al.*, 1986). Andersen (1970) used one animal, but did not report the ambient noise level, and did not explain how the thresholds were determined. The hearing of the harbor porpoise tested by Andersen, for frequencies above 8 kHz, was poor compared to that of other odontocetes (Thomas *et al.*, 1988), whereas anecdotal information at sea and in captivity suggests that harbor porpoise hearing is more sensitive. Also, the mismatch between the porpoise's own phonation and area of best hearing caused doubt about the validity of the audiogram reported by Andersen. Electrophysiological measures using auditory brain stem evoked potential responses (ABR) and click signals provide only a rough estimate of hearing thresholds because ABR is an onset phenomenon (Supin *et al.*, 1993). It is difficult to determine the appropriate sound pressure level values for a given response. Hearing thresholds are usually described in terms of the root mean square (rms) value of the sound pressure level (SPL) at the subject's threshold of hearing and for sounds of longer duration than the integration time. The onset feature of ABR makes it difficult to properly estimate the rms SPL associated with an ABR threshold. In addition, the ABR studies on underwater hearing in marine mammals are usually carried out in very small reverberating tanks, which make it difficult to control the actual SPL at the location of the animal's head.

[a]Electronic mail: r.kastelein@dolfinarium.nl
[b]Part of Grévin et Compagnie, France.

 0001-4966/2002/112(1)/334/11/$19.00

Hence, the underwater hearing abilities of a harbor porpoise needed reassessment as this could prove to be essential in developing a strategy to protect this marine mammal species. Fortunately, the opportunity arose to conduct hearing experiments after a stranded juvenile harbor porpoise had been raised at the Harderwijk Marine Mammal Park, The Netherlands. While investigation of the hearing sensitivity of the harbor porpoise could increase the fundamental knowledge of hearing abilities in cetaceans, this study also contributes to the ultimate goal of determining the optimal characteristics (e.g., frequency, SL, their number, distribution) of acoustic alarms devices mounted on gillnets that significantly reduce the bycatch of harbor porpoises.

II. MATERIALS AND METHODS

A. Subject

The study animal was a stranded male harbor porpoise (code PpSH047) which had been raised at the Netherlands Cetacean Research and Rehabilitation Center at the Harderwijk Marine Mammal Park from the age of approximately 8 months. During the experiment, the animal aged from 1.5 to 3 years, his body weight increased from 28 to 29 kg, his body length from 122 to 132 cm, and his girth anterior to the pectoral fins (at the auditory meatus) from 65 to 66 cm. Veterinary records showed that the animal had not been exposed to oto-toxic medication.

The animal received 1.2–2 kg of thawed fish (sprat, *Sprattus sprattus*, and herring, *Clupea harengus*) per day divided over 6 meals. The meal size during a hearing test session (first meal of the day) was disproportionally large (0.45 kg). The diet was supplemented by vitamins specially developed for marine mammals (Akwavit, Twilmij B. V., Stroe, The Netherlands). The animal had no previous experience with psychophysical testing.

B. Facility

The animal was kept in an indoor concrete oval pool [8.6 m (l)$\times$6.3 m (w), 1.2 m deep; Fig. 1] at the Research and Rehabilitation Center at the Harderwijk Marine Mammal Park, The Netherlands. Attention was paid to ensure a constant water level. Average water temperature was 19.5 °C, and average salinity was 2.2% NaCl. There was no current in the pool during the experiments, as the circulation pump (and the pump of the other pool in the building) was shut off 10 min before and during sessions. During the study period the study animal shared the pool with a 6-year-old female striped dolphin (*Stenella coeruleoalba*), a 1-year-old male harbor porpoise, or a 2-year-old female harbor porpoise. Nonstudy animals were kept quiet and fed at the opposite end of the pool during the experimental sessions in order to eliminate any distractions and interferences during a session. An adjacent room served as the observation and data collection laboratory, where all the controlling electronics were housed and where the equipment operator was seated during the experiments (Fig. 1). The operation of the equipment was not visible to the porpoise.

C. Signals and signal generation

A diagram of the signal generation system is shown in Fig. 2. Signals were produced by a wave form generator (Hewlett Packard, model 33120A). Each acoustic stimulus consisted of a narrow-band sinusoidally frequency-modulated (FM) signal (wobble) of 2.0 s duration. The signals had 150 ms rise and fall times to prevent abrupt signal onset and offset transients. The steady state portion of the signal thus was 1.7 s. The frequency modulation range of each stimulus signal was $\pm 1\%$ of the center frequency (i.e., the frequency around which the signal fluctuated symmetrically; see Table I), and the modulation frequency was 100 Hz. For example: if the signal's center frequency was 100 kHz, the frequency fluctuated 100 times per second (100 Hz) between 99 and 101 kHz ($\pm 1\%$). The study had started with pure tones, but this resulted in thresholds that varied by up to 15 dB between sessions. Repeated measurements of the SPL at the location of the porpoise's head indicated a difference in SPL of up to about 15 dB. These were suspected to have been caused by reflections of a long pure tone causing constructive and destructive interference of the signal. Therefore it was decided to use narrow-band FM signals. The advantage of using narrow-band FM signals is the reduction of propagation effects (multipath interferences) on the signals reaching the animal. However, narrow-band FM signals probably have a slightly higher arousal effect than pure tones, causing probably slightly lower thresholds. The use of narrow-band FM signals is therefore a trade-off: it provides a relatively stable SPL at the animal's head, but reduces comparability with previous studies on odontocete hearing.

A custom-built signal shaper and attenuator was used to control the amplitude of the signals. The sound pressure level at the porpoise's head while it was at the underwater listening station could be varied in 1dB steps. Before each session, the voltage output level of the system at the input of the transducer was verified with calibration levels (while the attenuator was at the same setting as during the calibrations) by using an oscilloscope (Dynatek, model 8120).

The lower frequency signals (250 Hz–32 kHz) were projected by an underwater low frequency piezoelectric transducer (Ocean Engineering Enterprise, USA, model DRS-8; 25 cm diameter) with an impedance matching transformer (Ocean Engineering Enterprise, USA). The 0.25, 0.5, and 1 kHz signals were amplified with a wide-band amplifier (Toellner, type TOE 7607). The 2–32 kHz signals needed no additional amplification. The higher frequency signals (32–180 kHz) were projected by a custom-built transducer consisting of a circular disk of 1–3 composite piezoelectric material (Material Systems Inc., Littleton, MA, U.S.A.), and an effective radiating aperture diameter of 4.5 cm. The thickness of the piezoelectric material was 0.64 cm. The piezoelectric element was a 6.4-cm-diam disk that was encapsulated in degassed polyurethane epoxy. The 32–180 kHz signals did not need amplification. The porpoise's hearing threshold for the 32 kHz signal was tested with both transducers. The sound propagation beam of each transducer was always aligned with the animal's body axis while it was at the station.

The porpoise's hearing sensitivity was measured at cen-

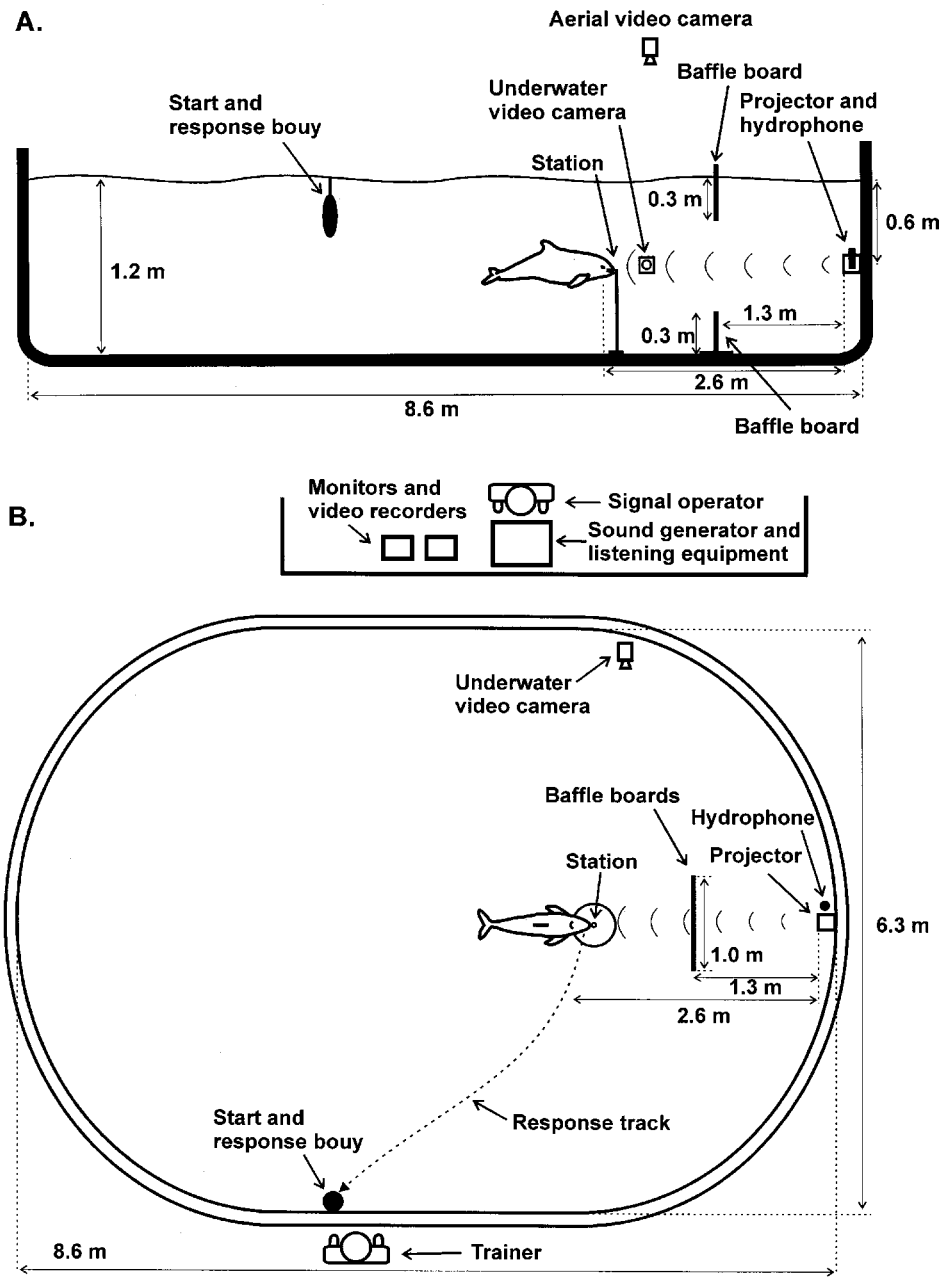


FIG. 1. The study area, showing the harbor porpoise in the correct position at the listening station: (a) sideview and (b) topview.

ter frequencies of 0.25, 0.5, 1, 2, 4, 8, 16, 32, 50, 64, 80, 100, 120, 130, 140, 150, 160, and 180 kHz. The low frequency cutoff of 250 Hz was determined by the limits of the sound production system, and was the lowest frequency the low frequency transducer could produce without distortions. Because the sonar signals of porpoises consist of narrow-band signals around 120 kHz (Møhl and Andersen, 1973; Verboom and Kastelein, 1995, 1997), the possibility of an acoustic fovea around 120 kHz was considered by using frequencies of 100, 130, 140, and 150 kHz. Originally, a one-octave frequency spacing for frequencies below 64 kHz was planned. However, the animal's sensitivity at 64 kHz was found to be lower than expected so that two additional frequencies (50 and 80 kHz) around 64 kHz were introduced. The high frequency cutoff of 180 kHz was determined by the limitations of the calibrating equipment to calibrate higher frequencies. The -3 dB beam width of the transducer at 130 kHz was approximately 15.6°. Since for a specific circular piston transducer the beam width is inversely proportional to frequency, the -3 dB beam width at 180 kHz should be approximately 11.3°. At a distance of 2 m, the -3 dB beam width will cover a circle having a diameter of 39 cm, which is larger than the head of the porpoise. The advantage of the transducers used in this study is that, at high frequencies, they are directional sound transmitters, causing less reflection from the sides of the pool. To reduce reflections from the water surface and the pool floor affecting the signal SPL at the animal's head, two baffle boards (6 mm thick, 30 cm $\times$100 cm aluminum plates covered with closed-cell neoprene) were placed perpendicular to the animal's axis, one breaking the water surface, and one at the pool floor (Fig. 1). The two boards were connected with nylon rope so they could be removed from the pool between sessions.

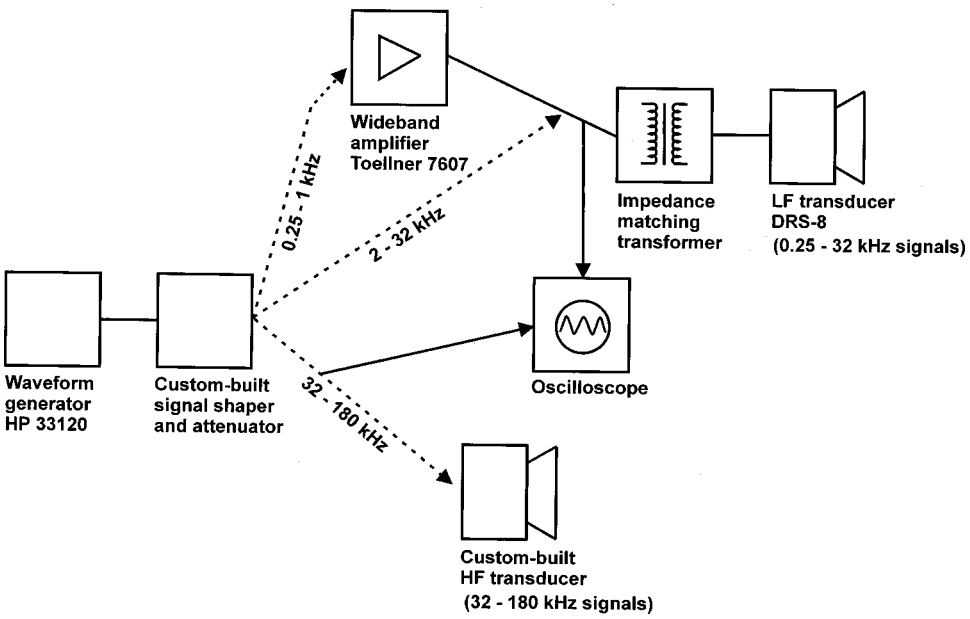


FIG. 2. Block diagram of the signal generation system used in the porpoise hearing study.

D. Signal calibration and monitoring

The signal SPLs (dB *re* 1 μPa, rms) for each test frequency were calibrated each month at the position of the porpoise's head when it was at the listening station during test sessions (2.6 m from the transducer). The porpoise was not at the station during the calibrations. The location of the porpoise's head while at the station relative to the transducer was always within a few cm in all six directions (he thus stayed within the 3 dB beamwidth of the transducer), and the animal did not adjust its position to a spot with a potentially higher signal SPL. The calibration equipment used for all signals consisted of a hydrophone (Brüel & Kjaer 8101; the calibration curve of this particular hydrophone showed that its frequency response was flat up to 100 kHz), a conditioning amplifier (Brüel & Kjaer, Nexus 2690), connected via a coaxial module (National Instruments, model BNC-2090) to a computer with an analog input/output card (National Instruments, PCI-MIO-16E-1, 12 bit resolution). The system was calibrated with a pistonphone (Brüel & Kjaer, 4223). For the calibration of signals above 100 kHz the frequency response of the measurement system was taken into account. To confirm this approach, a second hydrophone (B&K 8103) with a flat response ($+1$ dB/-2 dB) up to about 120 kHz was used with the frequency response of the measurement

TABLE I. The underwater 50% detection thresholds of a male harbor porpoise for 18 narrow-band FM signals, session threshold range, number of sessions, total number of reversals, and false alarm rate over all (signal present and signal absent) trials.

Center frequency (kHz)	FM range 1% of center frequency (kHz)	Mean 50% detection threshold (dB *re* 1 μPa)	Session threshold range (dB *re* 1 μPa)	Number of sessions (n)	Total No. of reversals	False alarm rate (%)
0.25	0.2475–0.2525	115	112–118	12	88	5
0.5	0.495–0.505	92	89–96	12	90	6
1	0.99–1.01	80	76–86	14	92	7
2	1.98–2.02	72	66–78	12	90	10
4	3.96–4.04	67	64–72	12	70	6
8	7.92–8.08	59	56–62	12	91	5
16	15.84–16.16	44	39–49	12	73	9
32	31.68–32.32	37[a]	28–42	15	103	9
50	49.5–50.5	36	33–39	12	99	8
64	63.36–64.64	46	40–51	12	75	3
80	79.2–80.8	37	36–40	12	90	5
100	99–101	32	29–35	12	80	9
120	118.8–121.2	33	31–37	12	75	9
130	128.7–131.3	35	28–40	12	73	5
140	138.6–141.4	36	32–41	12	83	5
150	148.5–151.5	60	57–63	12	83	4
160	158.4–161.6	91	87–97	12	91	10
180	178.2–181.8	106	97–111	12	84	7

[a]Based on the mean of 7 session thresholds measured with the LF transducer and 8 session thresholds determined with the HF transducer.

system taken into account. The values obtained with the B&K 8103 matched the results obtained with the B&K 8101.

The signals were digitized at a sample rate of 512 kHz in blocks of 0.2 s and fast Fourier transformed (FFT) into the frequency domain using a Hanning window. The highest peak in the spectrum was selected to determine the SPL and five consecutive 0.2 s time blocks were used to calculate the average SPL. Each month, the average SPL of each signal frequency was measured with the attenuator at the calibration setting (usually about 12 dB above the 50% detection threshold SPL). The analysis of the signals in the frequency domain was compared to rms analysis in the time domain, and the results matched. The linearity of the attenuator was also checked frequently.

Signals were analyzed with special attention to potential harmonics, especially with the low frequency, high amplitude signals (0.25, 0.5, and 1 kHz). These low frequency signals produced harmonics with energy well below the hearing thresholds obtained for those frequencies. The monthly SPL calibrations varied within 4 dB. The SPLs from the calibration nearest in time of the sessions were used to determine the session thresholds. Checks were often made for potential transients caused by pressing the signal button, both with and without the sound generator attached to the signal shaper/attenuator, or with the sound generator attached, but with the amplitude setting at 0. When the equipment was in the above-mentioned situations, the animal did not respond to the action of pressing the signal button. This was done at low and high amplitude settings of the sound generating system.

Before each session, the system was further verified by aurally monitoring the stimulus (usually at a higher amplitude than used during the session) via a hydrophone (LabForce 1BV) positioned directly adjacent to the transmitting transducer. The output of the monitoring hydrophone was connected to either an amplifier and loudspeaker for the frequencies up to 16 kHz, or to a bat-detector (Batbox III; Stag Electronics, Steyning, UK) for the signals with frequencies 32 kHz and above (maximum frequency possible was 120 kHz).

E. Background noise

Man-made noises in the vicinity were directly coupled into the pool. Therefore all indoor activities were stopped during sessions (nobody was allowed to move in the building). The water pumps in a nearby engine room were switched off 10 min before each session. The underwater background noise level was measured under the same conditions as during the study. Background noise in the pool was measured up to 8 kHz (at higher frequencies the ambient noise could not be measured, as the electronic noise of the recording system was higher than the ambient noise) by using the B&K 8101 hydrophone and the earlier described acoustic amplifier and conversion equipment. To allow comparison of the ambient levels and the hearing threshold levels, the recording and analysis methodology of the ambient levels was the same as that for the stimuli calibrations described previously. The recorded ambient signal was analyzed by FFT in ten blocks of 0.2 s. Of each block, the ambient levels of the tested frequencies were exported to a spreadsheet to calculate the average spectral level over 2 s (10 blocks). These levels are plotted in Fig. 4.

F. Experimental procedure

Training the porpoise for the hearing experiment took five months (August–December 1998). Operant conditioning using positive reinforcement was used for all training. A session began after the signal production equipment had been setup and the signal operator had set the frequency and the SPL for the first trial of the session. The amplitude in first trial of the first session of each frequency was set at about 20 dB above the threshold reported by Andersen (1970). A trial began with the animal stationed at the start and response buoy. When the trainer rang a bell, the animal swam to the listening station, which was the end of a 3-cm-diam water-filled PVC tube, and placed its head so that its auditory meatus was 2.6 m from the sound source, about 65 cm below the water surface (Fig. 1). He was trained to station with the tip of rostrum at the station and his body axis in line with the beam of the transducer. Because it was expected that the porpoise would have directional hearing (like in bottlenose dolphins; Au and Moore, 1984; Schlundt *et al.*, 2002), a maximum deviation in the porpoise's position of only 5° from the beam axis was accepted in all directions. Trials were canceled when the animal was not in the correct position. Due to the consistent behavior of the animal, no warm-up trials were conducted before the actual session began.

Signals were initiated following a random delay of 3 and 6 s after the porpoise stationed. If the animal detected the sound it left the station (go response) at any time during the 2 s signal duration and returned to the start and response buoy [Fig. 1(b)]. The signal operator told the trainer that the response was correct (a hit), after which the trainer gave a vocal signal and the porpoise received a fish. If the animal did not respond to the sound (a miss) the signal operator would tell the trainer that the animal's response was incorrect. The trainer would then signal the animal (by tapping on the side of the pool) that the trial had ended, thus calling him back to the start and response buoy. No reward was given. If the animal moved away before a signal was produced (a false alarm), the signal operator would notify the trainer to end the trial and not provide a reward to the animal.

For signal-absent trials, the signal operator told the trainer, after a random time period between 4 and 10 s after the porpoise had stationed, to end the trial by blowing a whistle. In case of a correct response (a correct rejection), the animal would return to the start buoy and receive a fish reward. If the porpoise left the station before the whistle was blown (a false alarm), the signal operator notified the trainer to end the trial and not reward the animal. The amount of fish reward for correct go and no-go trials was the same. After a correct response trial, the next trial would start as soon as the porpoise voluntarily returned to the start buoy.

A single frequency was tested during each session. A modified up/down staircase psychometric technique was used (Robinson and Watkins, 1973), a variance of the method of limits, which results in a 50% correct detection

threshold (Levitt, 1971). If the animal heard a signal and responded to it (a hit), the next signal presented was 4 dB less intense. If the animal did not hear a signal and remained at the station (a miss), the following signal levels were increased in 4 dB steps until the animal detected the signal again. The starting SPL of a session was set at 12–16 dB above the threshold found during that frequency's previous session. False alarms did not result in a change in signal amplitude for the next signal trial. A session usually consisted of 29 trials and lasted for about 15 min. Harbor porpoises have problems in doing more trials per session, in contrast to bottlenose dolphins *Tursiops truncatus* (Johnson, 1967) and false killer whales, *Pseudorca crassidens* (Thomas *et al.*, 1988). The attention span of small odontocete species seems shorter than those of larger species. Each session consisted of 50% signal-present and 50% signal-absent trials based on a pseudorandom series table (Gellermann, 1933), with the modification that the first trial in a session was always a signal-absent trial. To end in a positive way, the last trial was always followed by a correct response so the animal always received a reward after the last trial. Each session, one of 12 different data collection sheets with different Gellermann series was selected. Sessions with more than 15% false alarms were eliminated, because these usually coincided with restless behavior of the animal (swimming circles in the pool between trials and/or coming late to the start buoy). To avoid unintentional cueing, the trainer did not know before or during trials whether a signal was present or absent (double blind presentation). When the porpoise left the station, the operator observed the animal's behavior on two monitors in the cabin (Fig. 1), told the trainer whether or not to reward the porpoise, and recorded the animal's responses. Sometimes other behaviors were solicited from the porpoise between trials to occupy the animal during short periods of visible or audible disturbances outside the building. The order in which the frequencies were tested was randomized so that the effects of potential learning did not covary with test frequencies.

A switch in the porpoise's response from a detected signal (a hit) to a successive nondetected signal (a miss), and *vice versa*, is called a reversal. Amplitudes at which the animal reversed its response behavior were taken as data points. The mean 50% detection threshold per session was defined as the mean amplitude of the reversals in that session (usually 7 reversals per session; range 4–12). The overall mean threshold was determined by taking the mean of 12–15 session thresholds. Sessions were only used after the session threshold leveled off, which usually occurred after about 2–4 sessions with a particular frequency.

Data were collected between March 1999 and October 2000. Generally, one session was conducted daily (5 days/week) between 0830 and 0915 h (first feed of the day), so that the porpoise had not been fed for 15 h before a session, and when the park was still closed to visitors. When the park was closed between October 1999 and March 2000, two sessions were conducted on each of 5 mornings per week (between 0830–0915 and 1115–1200 h). No differences in average hearing thresholds were found between the 0830 and 1115 h sessions. In total 6300 trials (18 frequencies×12 sessions/frequency×29 trials/session) were analyzed.

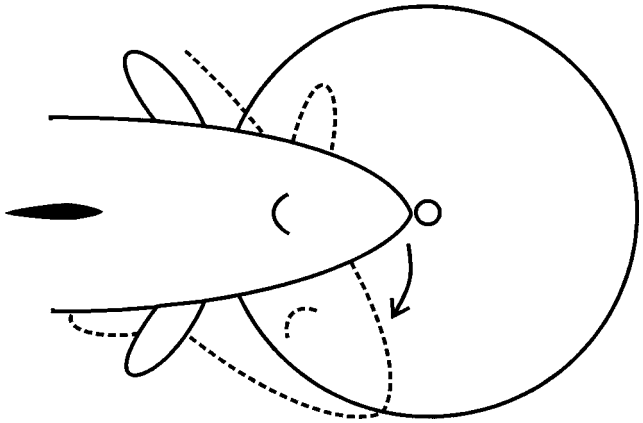


FIG. 3. The track of the porpoise's rostrum between the onset of the sound signal and the tip of the rostrum at the edge of the circle (a distance of 22 cm). The time this required is defined as the movement time.

G. Visual monitoring and image recording

To check the animal's position at the station and to gather data on his reactions and movement times to a stimulus, the animal's behavior was recorded by two video cameras. The images were visible in the signal operator's room. The porpoise was filmed from its left side by an underwater video camera (Mariscope, Micro, Kiel, Germany; Fig. 1). An aerial camera (Hapé, model CA 28) was hung from the ceiling just above the water surface and provided a top view of the animal. An infrared LED just in front of the lens, that was screened off so that it could not be seen by the animal, was filmed by the camera and showed when a sound was produced. Tests in which the LED was switched off showed that it was not a cue for the animal, as similar thresholds were reached as when the LED was operating (activated when the acoustic stimulus was presented). In addition, tests were done with only the LED being activated without the transducer being connected: the animal did not react to the LED's on switch. The images were recorded on videotape for later analysis.

H. Video analysis for movement time

From the images recorded by the aerial video camera, the movement time was measured to the nearest 40 ms (duration of one frame). The movement time was defined as the time between the onset of the signal (onset of LED) and the time the tip of the animal's rostrum was at the perimeter of a white 44-cm-diam white disk on the pool floor below the station (Fig. 3). The movement time was calculated for each signal trial of the following frequencies: 0.25, 0.50, 1, 2, 4, 8, 16, 32, 64, 100, 120, and 140 kHz. Per frequency a graph was made showing signal amplitude versus movement time and a trend line was drawn through all available data points (a minimum of ten movement time measurements per signal level was set as a minimum for a level to be included into the graph). Thereafter amplitudes were normalized to the 50% detection threshold of each particular frequency.

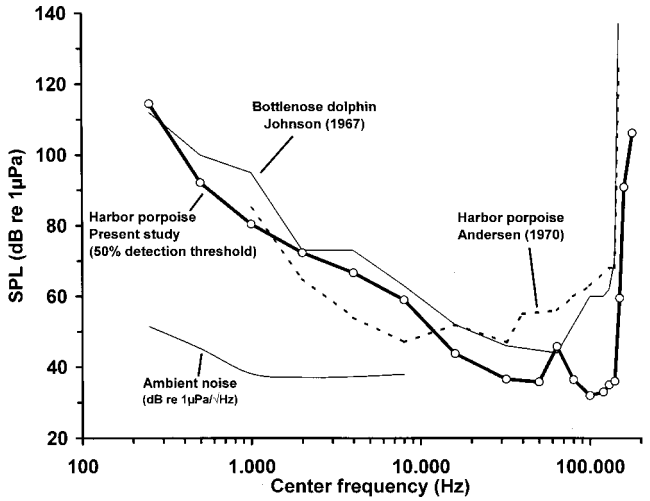


FIG. 4. The mean 50% detection thresholds in dB *re* 1 μPa (rms) for the tested narrow-band FM signals in the present study ($n=12-15$ mean session threshold per frequency, for details see Table I). Also shown is the audiogram determined by Andersen (1970) for one harbor porpoise (sample size per frequency threshold unknown, and definition of the threshold unknown), and the audiogram of an Atlantic bottlenose dolphin (Johnson, 1967). The spectral level (dB *re* 1 μPa/$\sqrt{}$Hz; note that this is a different unit than the one along the *Y* axis) of the ambient noise in the pool is shown up to 8 kHz.

III. RESULTS

A. Hearing sensitivity

The 50% detection thresholds for the 18 narrow-band FM signals of the harbor porpoise are listed in Table I. The resulting audiogram for this porpoise was U-shaped (Fig. 4), with hearing capabilities from 0.25 to 180 kHz (9.5 octaves). Maximum sensitivity (about 33 dB *re* 1 μPa) occurred between 100 and 140 kHz (for details see Table I). The range of most sensitive hearing (defined as 10 dB within maximum sensitivity) was from 16 to 140 kHz (3.1 octaves), with a reduced sensitivity around 64 kHz. The less sensitive hearing for this frequency was not a narrow-band phenomenon, as was shown by the slightly decreased sensitivity for the 50 and 80 kHz signals. The animal's hearing became less sensitive below 16 kHz and above 140 kHz. Sensitivity decreased by about 10 dB per octave below 16 kHz and fell sharply at a rate of 260 dB per octave above 140 kHz. The average false alarm response rate per frequency varied between 4% and 10% of all trials, and was on average over all frequencies 7% (Table I).

The 32 kHz average (over 7 sessions) threshold measured with the LF transducer was 36.4 ± 5.3 dB (*re* 1 μPa), while the average (over 8 sessions) threshold for the same frequency determined with the HF transducer was 36.8 ±4.1 dB. No significant difference was found between the mean thresholds (two-sample *T*-Test; $T=0.13$, degrees of freedom$=11$, $P=0.899$). Therefore all 15 session thresholds were used in the calculation of average 50% detection threshold for the 32 kHz signal. The match between the 32 kHz thresholds obtained with the two transducers suggests that the shape of the audiogram is not influenced by differences in transducer characteristics. The animal's sensitivity for each test frequency was stable over the two-year study period.

B. Movement times in relation to SPL

A negative relationship occurred between signal amplitude and movement time for all frequencies tested. When the SPL of a signal came closer to the animal's hearing threshold for that frequency, the animal would move more slowly than after louder signals (Fig. 5). The slope of the movement time was about -50 ms per dB above threshold. It usually took about 80 ms ($=$reaction time) between the onset of the signal and the beginning of the animal's head movement. In all cases, the animal reached the parameter of the circle with its rostrum during the signal presentation (2000 ms including the 150 ms rise and fall times). The variation of the 50% detection threshold between sessions varied per frequency. In some cases (1, 16, 32, and 64 kHz) the sample size for signals with SPLs below the 50% detection threshold was too low (<10) to be included into the graph. In general a session began about three 4-dB steps above the 50% detection threshold.

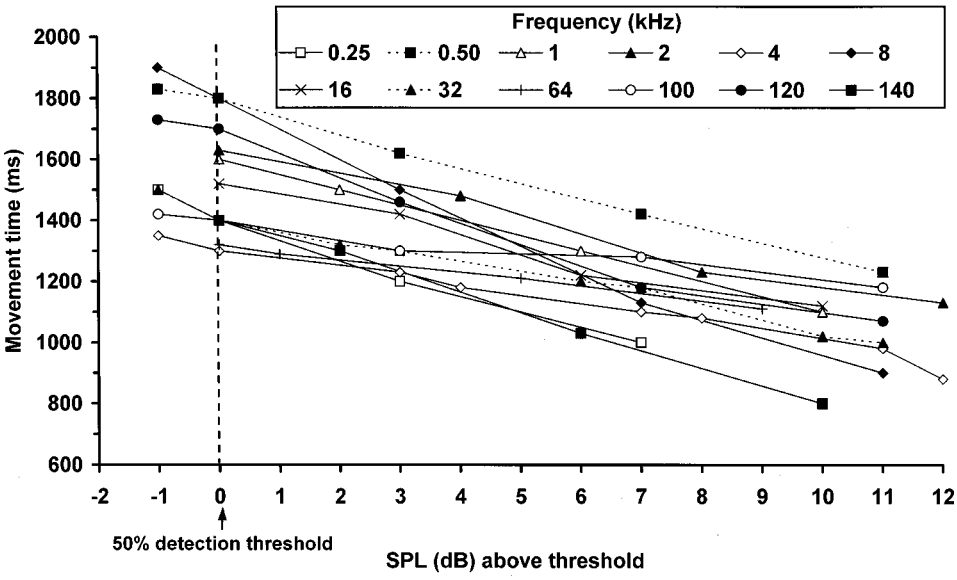


FIG. 5. The relationship between SPL relative to the 50% detection threshold and the average movement time of the porpoise for 12 narrow-band FM signals. The movement time decreases as the SPL increases. The sample size per data point increases as the SPL gets closer to the threshold, as a result of the up/down staircase method. In some cases (1, 16, 32, and 64 kHz) the sample size for signals with levels below the 50% detection threshold was too low (<10) to be included into the graph. In general a session began about three 4 dB steps above the 50% detection threshold.

IV. DISCUSSION AND CONCLUSIONS

A. Evaluation of the data

The comparatively low false alarm rate of the study animal is typical for marine mammals (Schusterman, 1974, 1976; Sauerland and Dehnhardt, 1998), but is also influenced by the way the animal was trained, and by the signal present/signal absent ratio. The porpoise probably only indicated the presence of a signal when it was confident of perceiving one. The present data, therefore, might represent a relatively conservative estimate of the auditory abilities of the harbor porpoise.

Two features of the method used influenced the 50% detection threshold: the pseudorandom Gellermann order in which the two trial types were presented, and the random time period between when the porpoise stationed and when the signal from the transducer or the whistle of the trainer were presented. Both these criteria reduced the possible anticipation by the animal, which may have influenced his response.

B. Comparison with previous harbor porpoise hearing studies

The audiogram of the present study deviates somewhat from the one made by Andersen (1970; Fig. 4). Between 2 and 8 kHz the hearing of the animal in the present study was less sensitive than that of the animal tested by Andersen. The thresholds in this frequency range found in the present study resemble more those found in other odontocetes (Johnson, 1967; Hall and Johnson, 1971; Jacobs and Hall, 1972; White *et al.*, 1978; Awbrey *et al.*, 1988; Thomas *et al.*, 1988; Wang *et al.*, 1992; Sauerland and Dehnhardt, 1998; Nachtigall *et al.*, 1995). However, above 8 kHz, the hearing of the animal in the present study was much more acute than that of the porpoise tested by Andersen. Age, sex, prior environment, individual differences, all could be factors contributing to the differences between the hearing sensitivity of the animal in the present study and that reported by Andersen (1970), as was shown to occur in bottlenose dolphins by Ridgway and Carder (1997). Differences in equipment and methodology could also have played a role. The results of the present study might be more representative of the hearing sensitivity of a young harbor porpoise with good hearing since the range of greatest hearing sensitivity found corresponds to the peak frequency of echolocation pulses of this species (120–130 kHz; Møhl and Andersen, 1973). A match between the frequency of sonar signals and area of best hearing is also found in the greater horseshoe bat (*Rhinolophus ferrumequinum*; Long, 1977). However, this match between the sonar signal frequency of peak energy and frequency range of highest hearing sensitivity is not found in the bottlenose dolphin (Johnson, 1967) and the false killer whale (Thomas *et al.*, 1988). In these species, the peak frequency of echolocation signals is dependent on the amplitude of the signal, and the bandwidth of the signals is wider than in the harbor porpoise and greater horseshoe bat. In general the frequency range of best hearing is around the average frequency of the echolocation signals.

Bibikov (1992) used auditory brainstem responses to test the hearing of a harbor porpoise that was contained in a very small tank (probably causing standing waves). He also found the lowest response threshold around 130 kHz. Popov *et al.* (1986) measured evoked potentials of the auditory cortex of a harbor porpoise also in a very small tank, and found the lowest evoked potential threshold curves within 120–130 kHz, but also found an additional sensitivity peak between 20 and 30 kHz. The latter peak was not found in the present psychophysical study, although the porpoise's hearing was fairly sensitive between 16 and 50 kHz. However, the auditory brainstem response method involves the use of short broadband signals so that the results are not very frequency-specific. In most cases, it is not clear to what frequency the auditory system is responding when it is given a short broadband acoustic stimulus.

Popov and Supin (1990) also found a decreased hearing sensitivity in the middle of audiograms; between 40 and 100 kHz in a harbor porpoise and between 30 and 60 kHz in an Amazon river dolphin (*Inia geoffrensis*). This supports the idea that the insensitivity phenomenon around 64 kHz found in the present study is a species, instead of an individual, characteristic.

The anatomical data of Ketten and Warzok (2000) suggest that the harbor porpoise should have a higher frequency of best hearing than was indicated by Andersen (1970). The high sensitivity region at 100 kHz observed in the present study is consistent with the measurement of ganglion cell density by Ketten. Ketten's preliminary data suggest an area of high ganglion cell density (over 10 000 cells/mm) located in the mid basal turn segment of the cochlea of the harbor porpoise, which would coincide with the 95–110 kHz region of the basilar membrane, and is in general agreement with data of the present study.

C. Comparison with hearing studies on other odontocetes

Underwater hearing thresholds have been determined in psychophysical tests for nine other odontocete species: The Atlantic bottlenose dolphin (Johnson, 1967; Fig. 4), killer whale, *Orcinus orca* (Hall and Johnson, 1971; Szymanski *et al.*, 1999), Amazon river dolphin, *Inia geoffrensis* (Jacobs and Hall, 1972), beluga whale, *Delphinapterus leucas* (White *et al.*, 1978; Awbrey *et al.*, 1988; Johnson, 1992; Klishin *et al.*, 2000), false killer whale (Thomas *et al.*, 1988), baji (Chinese river dolphin), *Lipotes vexillifer* (Wang *et al.*, 1992), tucuxi, *Sotalia fluviatilis guianensis* (Sauerland and Dehnhardt, 1998), Risso's dolphin, *Grampus griseus* (Nachtigall *et al.*, 1995), and Pacific white-sided dolphin, *Lagenorhynchus obliquidens* (Tremel *et al.*, 1998). In most of these studies the shape of the audiograms and the maximum sensitivities were fairly similar (maximum sensitivities were within about 10–15 dB). The upper frequency of hearing varied from about 90 to 150 kHz, except for the killer whale in the study of Hall and Johnson (1971).

The hearing sensitivity of the harbor porpoise in the present study is within the range of those of the other odontocetes up to around 32 kHz. However, above 32 kHz the

study animal's hearing is more acute and the upper frequency limit of hearing is higher than those of the other odontocetes.

D. Movement time

The movement time measured in the present study consists only for a very small percentage of the time difference between the LEDs on switch and the sound arriving at the animal. The time delay caused by the travel time of sound through the water (2.6 m distance/1500 m/s speed of sound) is 1.7 ms. This transmission time is negligible in relation to the movement times found in this study (850–1900 ms). These movement times are longer than the acoustic response times (145–448 ms) in bottlenose dolphins (Ridgway *et al.*, 1991), but this may be because the bottlenose dolphins only had to activate their vocalization system, whereas the porpoise in the present study had to put its entire body into action and move its rostrum 22 cm to the side (coping with the drag in water).

The closer the SPL of the signal to the 50% detection threshold, the slower the porpoise moved away from the station. In bottlenose dolphins, measuring acoustic responses to acoustic stimuli, Ridgway *et al.* (1991) also found that response times varied with stimulus amplitude, but also that it varied with stimulus duration, an aspect that was not tested in the present study. The increase in movement speed with increasing signal amplitude is fairly similar for the 12 analyzed frequencies. The slower movements with decreased SPL suggests that the animal uses some sort of likelihood criterion. It will continue to collect information about the signal, until it reaches a point where the likelihood that a signal was either present or absent exceeds some threshold value. In other words, the animal continues to listen until it is, say, 90% certain that it will make a correct decision. The higher movement times at lower SPLs suggest that the levels were reaching the hearing thresholds.

The reaction time (defined as the time between the onset of the signal and the start of the animal's movement) was fairly constant (around 80 ms) although the resolution of the analysis technique was low (one video frame is 40 ms). In humans, reaction time depends on the individual, the required behavior, and the stimulus (Henry, 1961; Carlson and Jensen, 1982). Comparison between studies is often difficult, because reaction time is influenced by the distance between the brain and the body parts that have to be activated (Davis, 1984).

E. Ecological significance and suggestions for future research

The frequency band of harbor porpoise echolocation signals lies within the frequency band of best hearing found in the present study. Thus, the hearing and echolocation signals of harbor porpoises are adapted for navigation and foraging in conditions where vision is limited or absent. Harbor porpoises are high latitude animals, living in an environment where biological noise is not high and broadband. The small size of harbor porpoise probably plays a major role in the use of higher frequencies for echolocation. For any frequency, the bigger the sound production organ, the narrower the beam. The harbor porpoise, being one of the smallest odontocetes, has a broad transmission beam (Au *et al.*, 1999), but the use of high frequency echolocation signals compensates for this as higher frequencies provide better directivity.

Relative to, for instance, the bottlenose dolphin, the sonar signals of the harbor porpoise have a lower SL (Au *et al.*, 1999), but the porpoise compensates for this to some degree by having lower hearing threshold levels around the frequency of peak energy. Despite its good hearing, the harbor porpoise's echolocation target detection abilities are poor (Kastelein *et al.*, 1999). Yet this is not surprising since the source level of the harbor porpoise is between 50 and 60 dB lower than that of the bottlenose dolphin (Au *et al.*, 1999) so that its better hearing sensitivity cannot compensate for such large differences in source levels.

The present study suggests that an acoustic alarm should have most energy between 100 and 140 kHz, the frequency range of best hearing of the harbor porpoise. Besides saving energy (such signals need to be less loud to be heard by porpoises than lower and higher frequency signals), such signals will be inaudible to most other marine organisms.

To be able to estimate the distance at which harbor porpoises can hear each other, acoustic alarms on fishing nets, or vessels under varying ambient noise conditions, additional information is needed. To understand fully how the harbor porpoise is adapted acoustically to its environment, information needs to be obtained on how it hears in the presence of masking noise (critical ratios, critical bands), how it hears sounds of different duration, and how well it spatially resolves sounds coming from different directions (directivity index).

ACKNOWLEDGMENTS

We thank Nicole Schooneman, Carolien Staal, Jolanda Meerbeek, Lauro Marcenaro, Joyce Borrias and Muriel Drenthe for their help with training and data collection, Pim Dorrestein for technical assistance, and Rob Triesscheijn for making the graphs. We thank Søren Andersen for his comments on the discussion and for additional information on his 1970 study. We also thank Peter van de Sman, Wim Verboom (TNO, Institute of Applied Physics, Delft, The Netherlands), Patrick Moore (Naval Command, Control and Ocean Surveillance Center, RDT&E Division-NRaD, San Diego), Jack Terhune (University of New Brunswick, Canada), Bertel Møhl (University of Aarhus, Denmark), Alexander Supin (Institute of Ecology and Evolution, Moscow, Russia), Nancy Vaughan (University of Bristol, UK), Paul Nachtigall (Hawaii Institute of Marine Biology, University of Hawaii), Lee Miller (Odense University, Denmark), and Jacob Tougaard (Odense University, Denmark), for their valuable constructive comments on this manuscript. This study was funded by the Harderwijk Marine Mammal Park, The Netherlands, and The North Sea Directorate [DNZ; through Wanda Zevenboom, Contract Nos. 76/316826; IBO 14.2 (1998) and 76/317581; IBO 14.1 (1999)] of the Directorate-General of the Netherlands Ministry of Transport, Public Works and Water Management (RWS). The porpoise's training and testing were conducted under authorization of the Netherlands Ministry of Agriculture, Nature Management

and Fisheries, Department of Nature Management. Endangered Species Permit No. FEF27 06/2/98/0184. We dedicate this study to Søren Andersen for his pioneer work on harbor porpoise acoustics.

Andersen, S. (**1970**). ''Auditory sensitivity of the harbour porpoise, *Phocoena phocoena*,'' in *Investigations on Cetacea*, edited by G. Pilleri (Institute for Brain Research, Bern), Vol. 2, pp. 255–259.

Au, W. W. L., Kastelein, R. A., Rippe, T., and Schooneman, N. M. (**1999**). ''Transmission beam pattern and echolocation signals of a harbor porpoise (*Phocoena phocoena*),'' J. Acoust. Soc. Am. **196**, 3699–3705.

Au, W. W. L., and Moore, P. W. B. (**1984**). ''Receiving beam patterns and directivity indices of the Atlantic bottlenose dolphin *Tursiops truncatus*,'' J. Acoust. Soc. Am. **75**, 255–262.

Awbrey, F. T., Thomas, J. A., and Kastelein, R. A. (**1988**). ''Low-frequency underwater hearing sensitivity in belugas, *Delphinapterus leucas*,'' J. Acoust. Soc. Am. **84**, 2273–2275.

Bibikov, N. G. (**1992**). ''Auditory brainstem responses in the Harbor porpoise (*Phocoena phocoena*),'' in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Ya Supin (Plenum, New York), pp. 197–211.

Carlson, J. S., and Jensen, C. M. (**1982**). ''Reaction time, movement time, and intelligence: A replication and extension,'' Intelligence **6**, 265–274.

Culik, B. M., Koschinski, S., Tregenza, N., and Ellis, G. (**2000**). ''Reactions of harbour porpoises (*Phocoena phocoena*) and herring (*Clupea harengus*) to acoustic alarms,'' Mar. Ecol.: Prog. Ser. **211**, 255–260.

Davis, M. (**1984**). ''The mammalian startle response,'' in *Neural Mechanisms of Startle Behavior*, edited by R. C. Eaton (Plenum, New York), pp. 287–351.

Gaskin, D. E. (**1992**). ''Status of the Harbour Porpoise, *Phocoena phocoena*, in Canada,'' Can. Field Naturalist **106**, 36–54.

Gearin, P. J., Gosho, M. E., Laake, J. L., Cooke, L., DeLong, R. L., and Hughes, K. M. (**2000**). ''Experimental testing of acoustic alarms (pingers) to reduce bycatch of harbour porpoise, *Phocoena phocoena*, in the state of Washington,'' J. Cetacean Research and Management **2**, 1–9.

Gellermann, L. W. (**1933**). ''Chance orders of alternating stimuli in visual discrimination experiments,'' J. Gen. Psychol. **42**, 206–208.

Hall, J. D., and Johnson, C. S. (**1971**). ''Auditory thresholds of a killer whale,'' J. Acoust. Soc. Am. **51**, 515–517.

Henry, F. M. (**1961**). ''Reaction time-movement time correlations,'' Percept. Mot. Skills **12**, 63–66.

Jacobs, D. W., and Hall, J. D. (**1972**). ''Auditory thresholds of a fresh water dolphin, *Inia geoffrensis* Blainville,'' J. Acoust. Soc. Am. **51**, 530–533.

Johnson, S. C. (**1967**). ''Sound detection thresholds in marine mammals,'' in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, New York), Vol. 2, pp. 247–260.

Johnson, S. C. (**1971**). ''Auditory masking of one pure tone by another in the bottlenose porpoise,'' J. Acoust. Soc. Am. **49**, 1317–1318.

Johnson, S. C. (**1992**). ''Detection of tone glides by the beluga whale,'' in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Ya. Supin (Plenum, New York), pp. 241–247.

Kastelein, R. A., Au, W. W. L., Rippe, H. T., and Schooneman, N. M. (**1999**). ''Target detection by an echolocating harbor porpoise (*Phocoena phocoena*),'' J. Acoust. Soc. Am. **105**, 2493–2498.

Kastelein, R. A., de Haan, D., Staal, C., Schooneman, N. M., and Vaughan, N. (**2001**). ''The influence of three acoustic alarms on the behaviour of harbour porpoises (*Phocoena phocoena*) in a floating pen,'' Marine Environ. Res. **52**, 351–371.

Kastelein, R. A., Goodson, A. D., Lien, J., and de Haan, D. (**1995**). ''The effects of acoustic alarms on Harbour porpoise (*Phocoena phocoena*) behaviour,'' in *Harbour Porpoises, Laboratory Studies to Reduce Bycatch*, edited by P. E. Nachtigall, J. Lien, W. W. L. Au, and A. J. Read (De Spil, Woerden, The Netherlands), pp. 157–167.

Kastelein, R. A., de Haan, D., Goodson, A. D., Staal, C., and Vaughan, N. (**1997**). ''The effects of various sounds on a harbour porpoise (*Phocoena phocoena*),'' in *The Biology of the Harbour Porpoise*, edited by P. E. Nachtigall, J. Lien, W. W. L. Au, and A. J. Read (De Spil, Woerden, The Netherlands), pp. 367–383.

Kastelein, R. A., Rippe, H. T., Vaughan, N., Schooneman, N. M., Verboom, W. C., and de Haan, D. (**2000**). ''The effect of acoustic alarms on the behavior of harbor porpoises (*Phocoena phocoena*) in a floating pen,'' Marine Mammal Sci. **16**, 46–64.

Ketten, D. R. (private communication).

Ketten, D. R., and Wartzok, D. (**2000**). ''Three-dimensional reconstructions of the dolphin ear,'' in *Sensory Abilities of Cetaceans*, edited by J. Thomas and R. Kastelein (Plenum, New York), pp. 81–105.

Klishin, V. O., Popov, V. V., and Supin, A. Ya. (**2000**). ''Hearing capabilities of a beluga whale, *Delphinapterus leucas*,'' Aquat. Mammals **26**, 212–228.

Kraus, S., Read, A. J., Solow, A., Baldwin, K., Spradlin, T., Williamson, J., and Anderson, E. (**1997**). ''Acoustic alarms reduce porpoise mortality,'' Nature (London) **388**, 525.

Laake, J., Rugh, D., and Baraff, L. (**1998**). ''Observations of harbor porpoise in the vicinity of acoustic alarms on a set gill net,'' NOAA Technical Memorandum NMFS-AFSC-84, U.S. Dept. of Commerce. 40 pp.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Lien, J., Hood, C., Pittman, D., Ruel, P., Borggaard, D., Chisholm, C., Wiesner, L., Mahon, T., and Mitchell, D. (**1995**). ''Field tests of acoustic devices on groundfish gillnets: Assessment of effectiveness in reducing harbour porpoise by-catch,'' in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp. 349–364.

Long, G. R. (**1977**). ''Masked auditory thresholds from the bat, *Rhinolophus ferrumequinum*,'' J. Comp. Physiol. **116**, 247–255.

Møhl, B., and Andersen, S. (**1973**). ''Echolocation: High-frequency component in the click of the harbour porpoise (*Phocoena ph.* L.),'' J. Acoust. Soc. Am. **53**, 1368–1372.

Nachtigall, P. E., Au, W. W. L., Pawloski, J. L., and Moore, P. W. B. (**1995**). ''Risso's dolphin (*Grampus griseus*) hearing thresholds in Kaneohe Bay, Hawaii,'' in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden), pp. 49–53.

Northridge, S. P. (**1991**). ''An updated world review of interactions between marine mammals and fisheries,'' FAO Fish. Tech. Paper 251 (Suppl. 1), 58 pp.

Popov, V. V., Ladygina, T. F., and Supin, A. Ya. (**1986**). ''Evoked potentials of the auditory cortex of the porpoise, *Phocoena phocoena*,'' J. Comp. Physiol., A **158**, 705–711.

Popov, V. V., and Supin, A. Ya. (**1990**). ''Electrophysiological studies of hearing in some cetaceans and a manatee,'' in *Sensory Abilities of Cetaceans*, edited by J. Thomas and R. Kastelein (Plenum, New York), pp. 405–415.

Ridgway, S. H., and Carder, D. A. (**1997**). ''Hearing deficits measured in some *Tursiops truncatus*, and discovery of a deaf/mute dolphin,'' J. Acoust. Soc. Am. **101**, 590–594.

Ridgway, S. H., Carder, D. A., Kamolnick, P. L., Skaar, D. J., and Root, W. A. (**1991**). ''Acoustic response times (RTs) for *Tursiops truncatus*,'' J. Acoust. Soc. Am. **89**, 1967–1968.

Robinson, D. E., and Watkins, C. S. (**1973**). ''Psychophysical methods in modern psychoacoustics,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. 2, pp. 99–131.

Sauerland, M., and Dehnhardt, D. (**1998**). ''Underwater audiogram of a *Tucuxi* (*Sotalia fluviatilis guianensis*),'' J. Acoust. Soc. Am. **103**, 1199–1204.

Schlundt, C. E., Carder, D. A., and Ridgway, S. H. (**2002**). ''The effect of projector position on the hearing thresholds of dolphins (*Tursiops truncatus*) at 2, 8 and 12 kHz.'' in *Echolocation in Bats and Dolphins*, edited by J. Thomas, C. Moss, and M. Vater (The University of Chicago Press, Chicago).

Schusterman, R. J. (**1974**). ''Low false-alarm rates in signal detection by marine mammals,'' J. Acoust. Soc. Am. **55**, 845–848.

Schusterman, R. J. (**1976**). ''California sea lion underwater auditory detection and variation of reinforcement schedules,'' J. Acoust. Soc. Am. **59**, 997–1000.

Supin, A. Ya., Popov, V. V., and Klishin, V. O. (**1993**). ''ABR frequency tuning curves in dolphins,'' J. Comp. Physiol., A **173**, 649–656.

Szymanski, M. D., Bain, D. E., Kiehl, K., Pennington, S., Wong, S., and Henry, K. R. (**1999**). ''Killer whale (*Orcinus orca*) hearing: Auditory brainstem response and behavioral audiograms,'' J. Acoust. Soc. Am. **84**, 936–940.

Thomas, J. A., Chun, N., Au, W. W. L., and Pugh, K. (**1988**). ''Underwater audiogram of a false killer whale (*Pseudorca crassidens*),'' J. Acoust. Soc. Am. **84**, 936–940.

Tremel, D. P., Thomas, J. A., Ramirez, K. T., Dye, G. S., Bachman, W. A., Orban, A. N., and Grimm, K. K. (**1998**). ''Underwater hearing sensitivity

of a Pacific white-sided dolphin, *Lagenorhynchus obliquidens*,'' Aquatic Mammals **24**, 63–69.
Trippel, E. A., Strong, M. B., Terhune, J. M., and Conway, J. D. (**1999**). ''Mitigation of harbour porpoise (*Phocoena phocoena*) by-catch in the gillnet fishery in the lower Bay of Fundy,'' Can. J. Fish. Aquat. Sci. **56**, 113–123.
Verboom, W. C., and Kastelein, R. A. (**1995**). ''Acoustic signals by Harbour porpoises (*Phocoena phocoena*),'' in *Harbour Porpoises, Laboratory Studies to Reduce Bycatch*, edited by P. E. Nachtigall, J. Lien, W. W. L. Au, and A. J. Read (De Spil, Woerden, The Netherlands), pp. 1–39.
Verboom, W. C., and Kastelein, R. A. (**1997**). ''Structure of click train signals of harbour porpoises (*Phocoena phocoena*),'' in *The Biology of the Harbour Porpoise*, edited by A. J. Read, P. R. Wiepkema, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp. 343–362.
Wang, D., Wang, K., Xiao, Y., and Sheng, G. (**1992**). ''Auditory sensitivity of a Chinese river dolphin, *Lipotes vexillifer*,'' in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Ya. Supin (Plenum, New York), pp. 213–221.
White, M. J. Jr., Norris, J., Ljungblad, D. K., Baron K., and di Sciara, G. (**1978**). ''Auditory thresholds of two beluga whales (*Delphinapterus leucas*),'' HSWRI Technical Report No. 78–109, Hubbs Marine Research Institute, San Diego, California.