CODEN: JASMAN

ISSN: 0001-4966

The Journal of the Acoustical Society of America

Vol. 110, No. 5, Pt. 1 November 2001

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 110, NO. 5, PT. 1, NOVEMBER 2001

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 110, NO. 5, PT. 1, NOVEMBER 2001

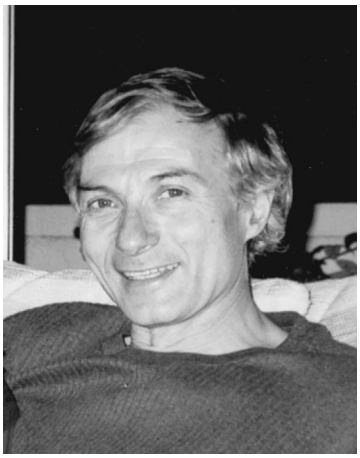
CONTENTS—*Continued from preceding page*

ACOUSTICAL NEWS—USA

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

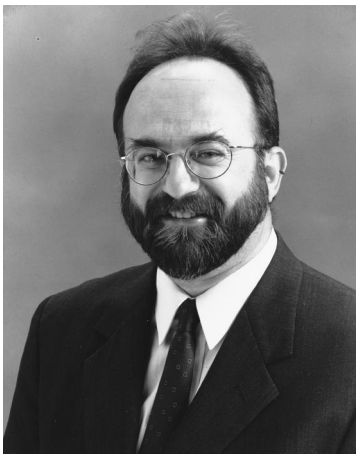
Editor's Note: *Readers of this Journal are encouraged to submit news items on awards, appointments, and other activities about themselves or their colleagues. Deadline dates for news items and notices are 2 months prior to publication.*

New Fellows of the Acoustical Society of America



Pierre L. Divenyi—For contributions to psychological acoustics.

Birger Kollmeier—For contributions to psychoacoustics and applied audiology.



Ahmet Selamet—For contributions to the acoustics of mufflers and silencers.

Michael Taroudakis—For contributions to geoacoustic inversion and propagation modeling.

 0001-4966/2001/110(5)/2237/2/$18.00

Jan D. Achenbach receives William Prager Medal

ASA Fellow Jan D. Achenbach, Director, Center for Quality Engineering and Failure Prevention at Northwestern University, was awarded the 2001 William Prager Medal in Solid Mechanics by the Society of Engineering Science.

The William Prager Medal is awarded for outstanding research contributions in either theoretical or experimental Solid Mechanics or both.

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.

2001

15–18 Nov. American Speech Language Hearing Association Convention, New Orleans, LA [American Speech-Language-Hearing Association, 10801 Rockville Pike, Rockville, MD 20852; Tel. 888-321-ASHA; E-mail: convention@asha.org; WWW: professional.asha.org/convention/abstracts/welcome.asp].

3–7 Dec. 142nd Meeting of the Acoustical Society of America, Ft. Lauderdale, FL [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]. Deadline for submission of abstracts: 3 August 2001.

2002

21–23 Feb. National Hearing Conservation Association Annual Conference, Dallas, TX [NHCA, 9101 E. Kenyon Ave., Ste. 3000, Denver, CO 80237; Tel: 303-224-9022; Fax: 303-770-1812; E-mail: nhca@gwami.com; WWW: www.hearingconservation.org/index.html].

10–13 March Annual Meeting of American Institute for Ultrasound in Medicine, Nashville, TN [American Institute of Ultrasound in Medicine, 14750 Sweitzer Lane, Suite 100, Laurel, MD 20707-5906; Tel: 301-498-4100 or 800-638-5352; Fax: 301-498-4450; E-mail: conv_edu@aium.org; WWW: www.aium.org].

3–7 June 143rd Meeting of the Acoustical Society of America, Pittsburgh, PA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org]. Deadline for receipt of abstracts: 1 February 2002

2–6 Dec. 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.html]. Deadline for receipt of abstracts: 28 June 2002

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. Price: ASA members $40 (paperbound); Nonmembers $90 (clothbound).

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ALIREZA DIBAZAR, *Dept. of BioMed Engr., Univ. of Southern California, Los Angeles, California 90089*
IBRAHIM M. HALLAJ, *Wolf Greenfield & Sacks P.C., 600 Atlantic Avenue, Boston, Massachusetts 02210*
HASSAN NAMARVAR, *Dept. of BioMed Engr., Univ. of Southern California, Los Angeles, California 90089*
DAVID PREVES, *4 Deerfield Drive, Princeton Junction, New Jersey 08550*
DANIEL RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
CARL J. ROSENBERG, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*

6,210,336

43.35.Ae DAMPING CUSHION FOR ULTRASOUND PROBES

Paul Sverre Fredriksen, assignor to G. E. Vingmed Ultrasound A/S
3 April 2001 (Class 600/459); filed 30 December 1998

This probe consists of a base and an ultrasonic transducer housed inside the base. An intermediary damping member secured to the base includes an interior cavity, which contains an acoustic coupling fluid. The probe may be optionally secured to a flexible intermediate buffer. In one version of the device, a coupling member secured to the base also includes an interior cavity filled with an acoustic damping fluid. The coupling features a malleable contacting surface to be pressed against an anatomical structure during use. Another embodiment of the device includes a transducer housing, an ultrasonic transducer mounted inside the housing, and a flexible cushioned pad mounted directly on the periphery of the housing. Yet another design includes a pliable damping member with an interior fluid cavity and a structure for attaching the damping member to an ultrasonic probe.—DRR

6,210,337

43.35.Sx ULTRASONIC ENDOSCOPIC PROBE

Paul T. Dunham *et al.*, assignors to ATL Ultrasound, Incorporated
3 April 2001 (Class 600/462); filed in the United Kingdom 7 June 1995

The ultrasonic endoscopic probe described in this patent features an articulating digital tip where an ultrasonic transducer is located. The articulating section of the probe can be locked into an articulated position by an external lock control. The user can vary the locking force so that the moveable section can be maneuvered and locked into position by a specific force. The articulating section is controlled by the adjustment of cable tension, thereby defining the range of articulation. The articulating section consists of alternating pivot rings with intervening polymeric pivot beads, which provide repeatable, smooth articulation. The ultrasonic transducer rotates to guide the acoustic scanning plane during use. A sliding membrane between the transducer and its acoustic window permits the transducer to rotate smoothly without sticking.—DRR

6,216,091

43.35.Ae ULTRASONIC MEASUREMENT SYSTEM WITH MOLECULAR WEIGHT DETERMINATION

Robert H. Hammond, assignor to Panametrics, Incorporated
10 April 2001 (Class 702/23); filed 25 September 1998

This device takes advantage of the well-known fact that the speed of sound in a gas depends on the average molecular weight. The ultrasonic measurement system described in the patent performs a signal path measurement by directing ultrasonic signals through a gaseous material in a conduit and processes the detected signals to determine the sound speed and to derive the average molecular weight of the unknown hydrocarbon mixture present in the material. The processor includes stored tables of critical constants of hydrocarbon mixtures as a function of the average molecular weight of the mixture and is configured to iteratively set a hypothetical molecular weight, determine the associated critical properties, and compute a predicted sound speed. If the two speeds differ, a new weight is assumed and the procedure is repeated until the predicted sound speed matches the measured speed, thus indicating that the assumed estimate is the correct average molecular weight.—DRR

6,210,356

43.35.Sx ULTRASOUND ASSEMBLY FOR USE WITH A CATHETER

James R. Anderson and Gary Lichttenegger, assignors to Ekos Corporation
3 April 2001 (Class 604/22); filed 5 August 1998

The subject device represents a dual system that contains an ultrasonic transducer for delivering ultrasound energy to a specific location within the body and also allows a variety of media to be delivered to treatment sites within the body. The device attempts to minimize the deleterious effects of ultrasound on the treatment sites. The catheter includes an elongated body with an exterior surface and an ultrasound transducer of specified length. A support member supports the ultrasound transducer at the exterior surface of the body. The chamber thus formed reduces transmission of ultrasound energy from the transducers into the elongated body along the length of the ultrasound transducer.—DRR

5,991,239

43.35.Wa CONFOCAL ACOUSTIC FORCE GENERATOR

Mostafa Fatemi-Booshehri and James F. Greenleaf, assignors to Mayo Foundation for Medical Education and Research
23 November 1999 (Class 367/164); filed 2 December 1996

The patent Background section is a well-written summary of medical ultrasonic imaging and focused ultrasonic force techniques. The invention is a multiple-element transducer capable of generating two or more high-frequency sound beams which intersect at a given location. The beams may differ in frequency, or one beam may be amplitude modulated.—GLA

6,206,831

43.35.Wa ULTRASOUND-GUIDED ABLATION CATHETER AND METHODS OF USE

Veijo T. Suorsa *et al.*, assignors to SciMed Life Systems, Incorporated
27 March 2001 (Class 600/439); filed 6 January 1999

Ablation is a process of breaking down materials by the application of heat. The device described in the patent provides ultrasound ablation catheters. In one embodiment, a tissue ablation apparatus includes a flexible elongated body having proximal and distal ends. Several spaced electrodes are attached to the flexible body near the distal end. A number of transducer elements are located between some of the electrodes. These transducers help the physician establish whether or not the ablation elements are in contact with the tissue to be ablated.—DRR

6,206,842

43.35.Wa ULTRASONIC OPERATION DEVICE

Lily Chen Tu and Hosheng Tu, both of Tustin, California
27 March 2001 (Class 601/2); filed 3 August 1998

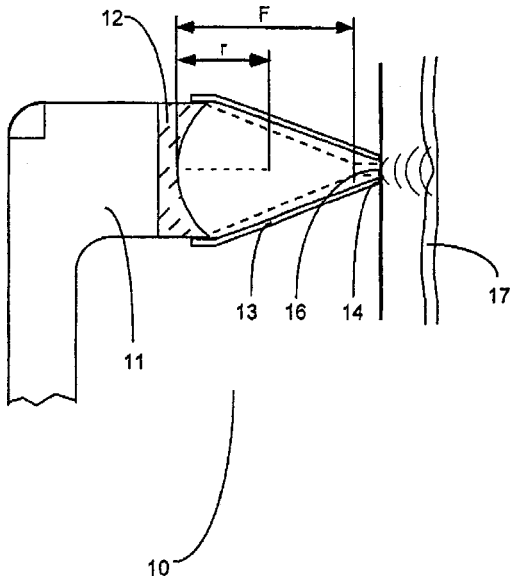
This device is designed to ablate cellular tissues using an ultrasonic beam aimed at the affected areas. Treatment of hemorrhoids or airway obstructions can be achieved by reducing the mass of cellular tissues. The device consists of an elongated tubular shaft housing at least one ultrasonic transducer and a vibration generator at the distal end of the tubular element to effect the ablation.—DRR

6,206,843

43.35.Wa ULTRASOUND SYSTEM AND METHODS UTILIZING SAME

Yoni Iger *et al.*, assignors to Ultra Cure Limited
27 March 2001 (Class 601/2); filed 28 January 1999

The device described by this patent is an ultrasonic system designed to exert a controlled acoustic pressure on an organ, for example, to empty a blood vessel or to stop leakage from an injured blood vessel. Rather than immersing the treated patient in a liquid tank or immersing the whole ultrasonic system in a liquid-filled container, the ultrasound-generating element is coupled to a container holding a liquid medium. The focal point of the ultrasonic beam is created inside the container and the pressure wave propagates from the container in the desired direction.—DRR

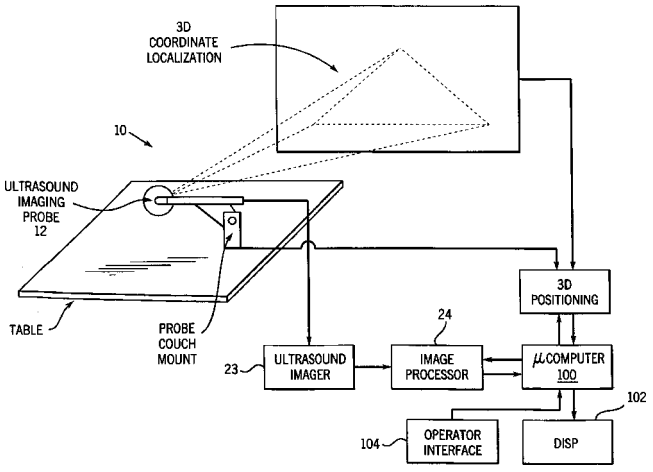


6,208,883

43.35.Wa ULTRASOUND LOCALIZATION AND IMAGE FUSION FOR THE TREATMENT OF PROSTATE CANCER

Edward I. Holupka *et al.*, assignors to Associates of the Joint Center for Radiation Therapy, Incorporated
27 March 2001 (Class 600/407); filed 26 July 1995

The patent covers a device to localize a patient's prostate using combined ultrasonics and X-rays and to treat the affected areas by X-rays. The system uses an ultrasound probe to generate an image of the prostate using fiducial end points as landmarks. The system also includes a cathode ray apparatus to generate X-ray simulator images of the anatomical region in the vicinity of the prostate and includes software to transform the ultrasound image into the simulator image. This enables accurate localization of the prostate relative to other anatomical features, thereby guiding the radiation therapy to be carried out on the prostate.—DRR



6,210,335

43.35.Wa ACOUSTIC FLASH TO INCREASE PENETRATION

Steven C. Miller, assignor to General Electric Company
3 April 2001 (Class 600/454); filed 8 December 1999

The aim of this device is to control the output of an ultrasonic transducer so as to permit greater penetration into anatomical structures while keeping the signals within the FDA mechanical limit and with sufficient dwell times between HAO frames to remain below the FDA temporal limits. This is done by transmitting two sets of ultrasound waves toward the subject and receiving the reflected signals, preferably using a transducer array. The second ultrasound transmission uses a lower power level than the first. The ratio of the durations of the first to the second reflected signals indicates the tissue compliance, mechanical index, temperature, and temporal-average spatial intensity peak, while enabling penetration of the ultrasound waves during the first transmission. The signals are processed to yield color flow data for display as a color flow image on a monitor.—DRR

6,215,226

43.35.Yb ULTRASONIC LIQUID GAUGING SYSTEM

Scott Robert Durkee, assignor to Simmonds Precision Products, Incorporated
10 April 2001 (Class 310/319); filed 23 December 1997

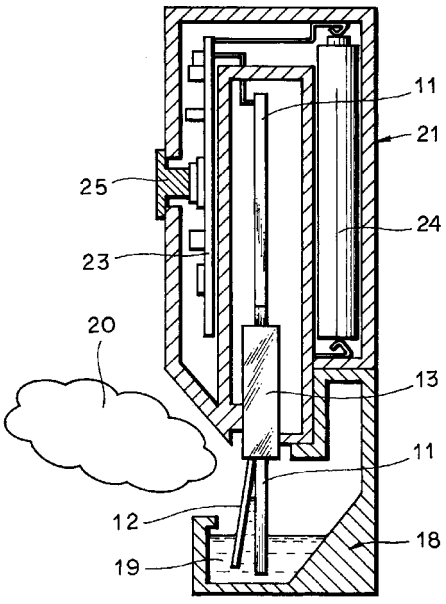
A liquid level gauge using pulse-echo and processing techniques is shown. The system uses several acoustic targets and the surface or liquid interfaces to obtain the echoes. Also described is a way for obtaining level data in thermally stratified media using parabolic approximations.—IMH

5,996,903

43.35.Zc ATOMIZER AND ATOMIZING METHOD UTILIZING SURFACE ACOUSTIC WAVE

Kei Asai and Takanobu Yamauchi, assignors to OMRON Corporation
7 December 1999 (Class 239/4); filed in Japan 7 August 1995

In this age of motorized towel dispensers and electric toothbrushes one would not expect asthma sufferers to rely on manually operated atomizers. However, small, battery-operated devices are not that easy to design. In this invention, vibrating paddle **11** generates surface acoustic waves in liquid **19**.



A controlled amount of liquid is drawn into a small cavity formed between the paddle and perforated plate **12**, where it is atomized and sprayed outward **20**. In comparison with prior art, the device is said to provide more uniform atomization from a battery of small power.—GLA

6,209,555

43.35.Zc SUBSTRATE CASSETTE FOR ULTRASONIC CLEANING

Kenneth C. Struven, assignor to Imtec Acculine, Incorporated
3 April 2001 (Class 134/182); filed 27 April 1999

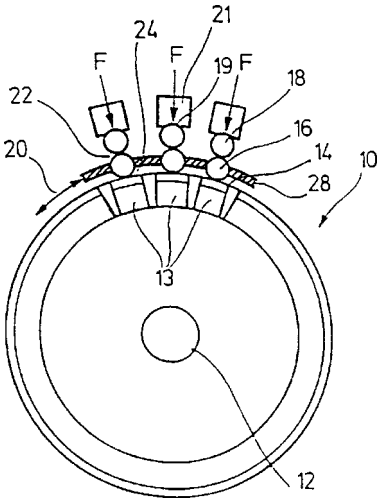
A megasonic cleaning chamber (cassette) is disclosed. The cassette has side holders and a lower holder for holding, e.g., semiconductor wafers. The holder material, or supports, is made of an impedance-matched material that permits the acoustic energy to propagate into the cleaning liquid.—IMH

6,213,377

43.35.Zc ULTRASONIC WELDING PROCESS AND ULTRASONIC WELDING DEVICE AND WELD PRODUCED THEREBY

Martin Schwerdtle and Arno Altpeter, assignors to Robert Bosch GmbH
10 April 2001 (Class 228/110.1); filed in Germany 24 September 1994

An ultrasonic welder is described for welding multiple wires to a connection point. An example use is for making commutator connections in electric motors. The system uses ultrasonic vibration along a direction **20**,



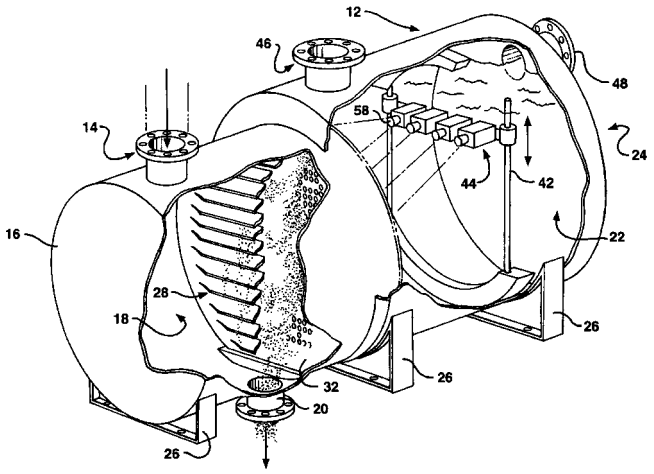
normal to the axis of the wires **16** and **18**, and normal to a contact force **F**.—IMH

6,221,255

43.35.Zc ULTRASOUND-ASSISTED FILTRATION SYSTEM

Achyut R. Vadoothker, Taylors, South Carolina
24 April 2001 (Class 210/739); filed 26 January 1998

In a system for filtering solid particles out of a fluid, an ultrasonic projector **58** blasts filtered particles out of the filter **32**. The particles, which remain on the upstream side of the filter, then settle due to gravity to the



bottom of the filter chamber, where they are removed through a drain **20**.—IMH

6,221,258

43.35.Zc METHOD AND APPARATUS FOR ACOUSTICALLY DRIVEN MEDIA FILTRATION

Donald L. Feke *et al.*, assignors to Case Western Reserve University
24 April 2001 (Class 210/748); filed 14 June 1996

A porous filter for filtering particles from a fluid is described. The particles may be of a size (e.g., 1–100 μm) smaller than the filter medium's pore size (e.g., 450 μm). The concept involves imposing a resonant acoustic field on the filter chamber that traps the small particles in the filter.—IMH

6,210,344

43.38.Ar METHOD AND APPARATUS FOR PASSIVE HEART RATE DETECTION

Dino Perin and Richard A. Riedel, assignors to UMM Electronics, Incorporated
3 April 2001 (Class 600/528); filed 24 March 1999

The device covered by this patent is a stethoscope modified to convert acoustic signals into an electronic digital output that can be displayed. In one embodiment the device consists of a hollow bell mounted on a diaphragm. A transducer element in the bell accepts sound transmitted through the diaphragm and converts this sound into digital form. A microprocessor performs a frequency analysis of the wave pattern data and sends the result to a digital output display.—DRR

6,044,160

43.38.Bs RESONANT TUNED, ULTRASONIC ELECTROSTATIC EMITTER

Elwood G. Norris, assignor to American Technology Corporation
28 March 2000 (Class 381/191); filed 13 January 1998

Tightly controlled ultrasonic beams can be produced by relatively small transducers. Two beams with a frequency difference of, say, 2 kHz would be expected to generate a virtual sound source at their point of intersection. If one beam is modulated, then audible voice and music can be reproduced. The idea is not new and several patents exist. In a parametric variant, only a single beam is used. The invention is an electrostatic transducer specifically designed for this application. An array of tuned cavities is located behind a single tensioned diaphragm. Negative air pressure biases the diaphragm, producing concave dimples which act as individual ultrasonic transducers.—GLA

6,175,636

43.38.Bs ELECTROSTATIC SPEAKER WITH MOVEABLE DIAPHRAGM EDGES

Elwood G. Norris and James J. Croft III, assignors to American Technology Corporation
16 January 2001 (Class 381/398); filed 26 June 1998

The diaphragm of a typical electrostatic loudspeaker is clamped around its perimeter. In response to electrostatic forces the diaphragm behaves like a drumhead. The volume of air displaced for a given maximum excursion is perhaps half that of an equivalent moving piston. In this invention only the top and bottom edges are clamped, so that when displaced from its rest position the diaphragm takes the form of a cylindrical surface. In theory, at least, low-frequency efficiency should be substantially improved.—GLA

6,081,064

43.38.Hz ACOUSTIC TRANSDUCER SYSTEM

Helmut Pfeiffer *et al.*, assignors to Endress+Hauser GmbH +Company
27 June 2000 (Class 310/334); filed in Germany 30 December 1997

The invention is intended for use in echo ranging systems. A flexural plate is driven at a frequency which produces well-defined, concentric, antinodal zones. Even-order and odd-order zones vibrate in opposing phase. However, mass rings are mounted to the rear of the plate at locations which reduce the amplitude of even-order zones, resulting in a beam of high directivity perpendicular to the plate.—GLA

6,128,395

43.38.Hz LOUDSPEAKER SYSTEM WITH CONTROLLED DIRECTIONAL SENSITIVITY

Gerard Hendrik Joseph De Vries, assignor to Duran B.V.
3 October 2000 (Class 381/387); filed in World IPO 8 November 1994

Suppose that you want to design a constant-directivity, wide-range loudspeaker array using any one of several known techniques. It turns out that logarithmic spacing is an efficient way to minimize the number of speakers. Even so, you will soon discover that practical loudspeakers are too large to fit the center-to-center spacing required at high frequencies. The inventor suggests that separate high- and low-frequency arrays be located side by side, just as Rudy Bozak did more than 25 years ago.—GLA

6,183,419

43.38.Hz MULTIPLEXED ARRAY TRANSDUCERS WITH IMPROVED FAR-FIELD PERFORMANCE

Douglas Glenn Wildes, assignor to General Electric Company
6 February 2001 (Class 600/447); filed 1 February 1999

Although the invention is aimed at ultrasound imaging, the technique could conceivably be applied to other transducer arrays. Controlling a large multi-row array to perform beamforming and focusing operations can be computationally simplified by multiplexing several cycles of successive subsets. The patent describes a sophisticated multiplexer that can perform a variety of operations using relatively few switches.—GLA

6,024,407

43.38.Ja VIBRATING PARTICLE MATERIAL FILLED FURNITURE

Byron C. Eakin, assignor to Somatron Corporation
15 February 2000 (Class 297/217.4); filed 10 April 1998

Numerous earlier patents have described ways in which audio signals can be used to vibrate dance floors and chairs. Also, simple inertia motors can be installed in a chair to provide a relaxing massage. The invention combines the two categories by employing acoustic energy from an embedded loudspeaker to vibrate the pellets used to stuff bean bag furniture.—GLA

6,095,279

43.38.Ja LOUDSPEAKER SYSTEM

Alan Brock Adamson, Scarborough, Ontario, Canada
1 August 2000 (Class 181/152); filed 31 July 1995

In 1992, Heil patented a simple, clever acoustic waveguide which generated a phase-coherent, line source when driven by a small piston or a conventional high-frequency driver. In some practical applications it would be desirable to have such a device take the form of a right-angle coupler and this is what the patent describes.—GLA

6,134,337

43.38.Ja LOUDSPEAKER

Takashi Isaka, assignor to Bell Tech Company, Limited; Foster Electric Company, Limited
17 October 2000 (Class 381/423); filed in Japan 11 December 1996

Soft drinks are packaged in lightweight plastic bottles. The bottom of such a bottle is molded into a complicated, hyperbolic paraboloid shape to provide a rigid base. The invention is a loudspeaker cone employing similar geometry.—GLA

6,160,898

43.38.Ja SUSPENSION MOUNT FOR SOUND REPRODUCTION DEVICES ACCORDING TO THE FLEXURAL WAVE PRINCIPLE

Wolfgang Bachmann *et al.*, assignors to Nokia Technology GmbH
12 December 2000 (Class 381/425); filed in Germany 20 December 1997

A large, distributed mode loudspeaker (DML) can be freely suspended and driven by an inertia motor attached directly to the panel. Several patents describe this arrangement. For smaller units it may be advantageous to resiliently mount the panel in a frame or enclosure. The patent includes a discussion of practical considerations and describes an edge suspension that provides performance benefits yet is economical to manufacture.—GLA

6,163,613

43.38.Ja LOW-DISTORTION LOUDSPEAKER

Kenneth W. Cowans, Fullerton, California
19 December 2000 (Class 381/55); filed 26 June 1995

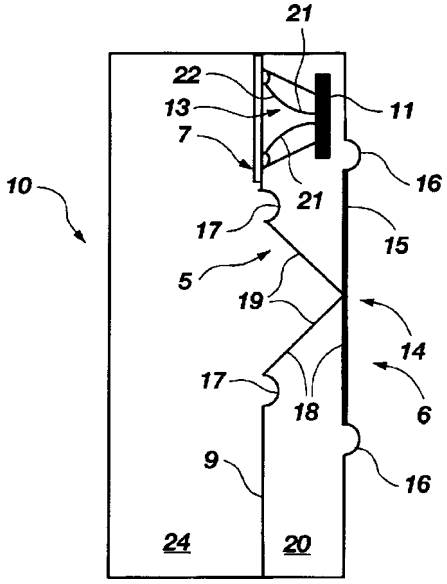
Direct radiator loudspeakers are a few percent efficient at best. Most of the signal power is actually used to heat the voice coil. In high-power loudspeakers, excess heat leads to all sorts of problems. The inventor has found a simple, elegant remedy. First, increase magnetic density in the voice coil gap to the highest practical value. Then add an external series resistor to restore the desired electrical system Q. Voila! Excess heat is now dissipated in the resistor, which can be located almost anywhere. The inventor has given this modest insight enough thought to expand it into 28 patent claims.—GLA

6,169,811

43.38.Ja BANDPASS LOUDSPEAKER SYSTEM

James J. Croft III, assignor to American Technology Corporation
2 January 2001 (Class 381/186); filed 2 March 1999

Loudspeaker **11** is mounted on partition **9** between chambers **20** and **24**. Passive radiator **5** is mounted on the same partition, but is mechanically coupled to larger passive radiator **14** on an exterior boundary of box **10**. The basic differential drive scheme was first described by Clarke in 1980 and a



similar ''Acoustic Lever'' was developed by Geddes in 1999, but these are both different from the arrangement shown. The patent includes more than two dozen variants, all intended to deliver bandpass response and all rigorously defined in 37 claims.—GLA

6,055,080

43.38.Kb OPTICAL MICROPHONE

Norbert Furstenau and Werner Jungbluth, assignors to Deutsche Forschungsanstalt fur Luft-und Raumfahrt e.V.
25 April 2000 (Class 359/151); filed in Germany 13 June 1996

The patent describes an interesting optical microphone design utilizing a light source, a fiber optic cable, and a dual-reflector interferometer. One reflector is the end face of the optical cable itself, which is mechanically coupled to the microphone diaphragm. According to the patent document, optical signals from the microphone can be transmitted more than a kilometer without any intermediate amplification.—GLA

6,154,551

43.38.Kb MICROPHONE HAVING LINEAR OPTICAL TRANSDUCERS

Anatoly Frenkel, Santa Fe, New Mexico
28 November 2000 (Class 381/172); filed 25 September 1998

The inventor states that optoelectronic technology has improved to the point that optical transducers offer significant advantages over electromagnetic or electrostatic transducers, particularly for microphones. These advantages include ''...higher resolution, higher signal-to-noise ratio, immunity to electromagnetic radiation, and greater linearity.'' In one embodiment, special beam-forming elements are used in conjunction with a dual-cell detector to

form a system whose output signal is independent of the intensity of the light. Several other interesting embodiments are also described, but there is no indication that any of them were actually built.—GLA

6,154,554

43.38.Kb MICROPHONE

Kazuhisa Kondo, assignor to Kabushiki Kaisha Audio-Technica
28 November 2000 (Class 381/355); filed in Japan 30 April 1996

Many conventional microphones use a damped resonance chamber in front of the diaphragm to extend and heighten high-frequency response. Unfortunately, the presence of damping material tends to roll off the high frequencies that are supposed to be extended. The invention employs two parallel acoustic cavities. One primarily sets the resonance frequency and the other controls the damping.—GLA

6,158,902

43.38.Kb BOUNDARY LAYER MICROPHONE

Raimund Staat, assignor to Sennheiser Electronic GmbH & Company KG
12 December 2000 (Class 391/357); filed in Germany 30 January 1997

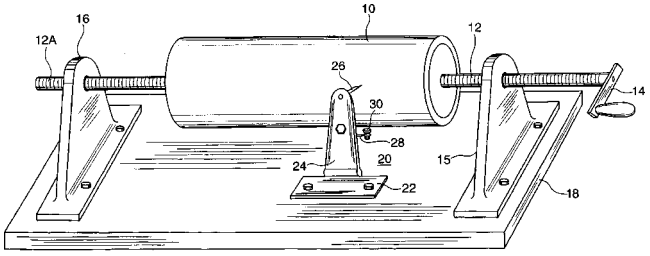
A cardioid microphone element is combined with a short front interference tube. The entire assembly is buried within a relatively thin, flat plate that can be set on a table top or attached to a large reflecting surface. The patent document includes a clear theoretical explanation along with response measurements that demonstrate improved performance in comparison with other directional boundary surface microphones.—GLA

6,185,179

43.38.Ne APPARATUS AND METHOD FOR RECORDING AND PLAYING BACK SOUND

Carl M. Mohrin, Hampton, New Jersey
6 February 2001 (Class 369/127); filed 31 August 1998

In Edison's original phonograph, sound waves vibrated a stylus which embossed a groove of varying depth in a rotating foil-covered cylinder. It would seem almost impossible to simplify this basic apparatus, and until



now no one has thought to try. But suppose that the foil-covered cylinder itself is flexible enough to vibrate in the presence of sound waves, and the stylus is fixed. Think about it.—GLA

6,154,639

43.38.Si HANDSFREE UNIT

Takuya Kanazumi *et al.*, assignors to Kabushiki Kaisha Toshiba
28 November 2000 (Class 455/74); filed in Japan 28 November 1997

Speakerphones have been commercially available for more than 40 years. Early designs relied on physical separation between the microphone and the loudspeaker to minimize acoustic feedback. In the following years, numerous analog circuits were developed to dynamically attenuate send and receive channels without seriously affecting two-way communication. According to this patent, the introduction of digital audio processing complicates the situation and can lead to unwanted side effects. The invention is a ''...cheap handsfree unit which can prevent the howling phenomenon...in both the analog and digital modes.''—GLA

6,195,437

43.38.Si METHOD AND APPARATUS FOR INDEPENDENT GAIN CONTROL OF A MICROPHONE AND SPEAKER FOR A SPEAKERPHONE MODE AND A NON-SPEAKERPHONE AUDIO MODE OF A COMPUTER SYSTEM

Mitchell A. Markow *et al.*, assignors to Compaq Computer Corporation
27 February 2001 (Class 381/93); filed 30 September 1997

This automatic microphone gain control system is intended for use in a computer where the microphone may be active either in speakerphone mode or in a non-speakerphone mode. The patent is not clear at all on just what the difference is supposed to be between these two modes in the case where a mic is being used for both modes. The speakerphone mode is said to use an additional DSP processor not used otherwise. There is no mention of a possible feedback cancellation function of that DSP. The patent merely acknowledges that the DSP performs a gain control function, which must be implemented in some other way for non-speakerphone mode.—DLR

6,212,275

43.38.Si TELEPHONE WITH AUTOMATIC PAUSE RESPONSIVE, NOISE REDUCTION MUTING AND METHOD

Akhteruzzaman, assignor to Lucent Technologies, Incorporated
3 April 2001 (Class 379/421); filed 30 June 1998

The patent describes a device to mute a telephone automatically when the local user pauses or stops speaking. Instead of half- or full-duplex operation, the system provides a signal comparator, which compares the speaker and the microphone signals. If the signals are the same, then it is determined that the local user is not speaking, and the microphone is muted to prevent the feedback common to speaker phone systems.—IMH

6,125,187

43.38.Tj HOWLING ELIMINATING APPARATUS

Mitsuru Hanajima *et al.*, assignors to Sony Corporation
26 September 2000 (Class 381/83); filed in Japan 20 October 1997

Detection and suppression of acoustic feedback (howling) in voice amplification systems has been a subject of concern for more than half a century. In recent years, inexpensive digital signal processing circuits have been developed to perform both functions automatically while the sound system is in use. The invention utilizes a two-step approach in which howling is first identified in a one-octave band and then subjected to narrow-band analysis and suppression.—GLA

6,137,887

43.38.Tj DIRECTIONAL MICROPHONE SYSTEM

Matthew G. Anderson, assignor to Shure Incorporated
24 October 2000 (Class 381/92); filed 16 September 1997

In 1984, Shure disclosed a clever automatic microphone mixer which used special dual-cardioid microphones. Only sounds arriving within a specified acceptance angle could gate a microphone. In this updated version,

electronic circuitry identifies which microphone best hears a given talker and gates only one microphone per talker.—GLA

6,122,381

43.38.Vk STEREOPHONIC SOUND SYSTEM

Martin Winterer, assignor to Micronas Interuetall GmbH
19 September 2000 (Class 381/17); filed in European Patent Office 17 May 1996

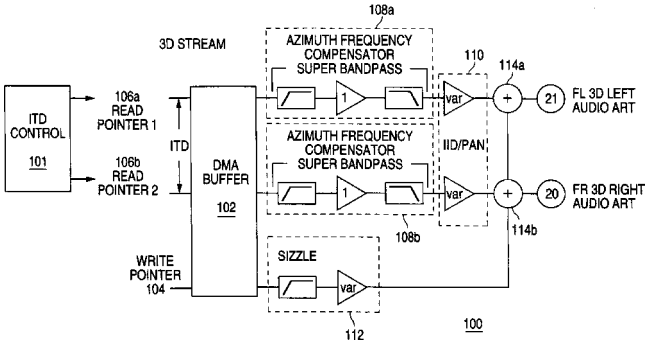
The left and right speakers housed in a conventional television set are spaced too closely for good stereo reproduction. Electronic sound field expansion can provide a more spacious stereo effect, but may enlarge the phantom center image at the same time. The invention processes left and right signals independently from the L+R center signal to minimize unwanted side effects.—GLA

6,178,245

43.38.Vk AUDIO SIGNAL GENERATOR TO EMULATE THREE-DIMENSIONAL AUDIO SIGNALS

David Thomas Starkey and Anthony Martin Sarain, assignors to National Semiconductor Corporation
23 January 2001 (Class 381/17); filed 12 April 2000

Virtual reality systems must generate convincing virtual sound sources at any location relative to the listener's head. At the present time, most such systems are based on head-related transfer functions. However, this approach is computationally intensive, even when various simplifying schemes are used. The invention makes use of only three easy-to-implement parameters to orient a virtual sound source along any specified axis.—GLA



6,181,800

43.38.Vk SYSTEM AND METHOD FOR INTERACTIVE APPROXIMATION OF A HEAD TRANSFER FUNCTION

J. Andrew Lambrecht, assignor to Advanced Micro Devices, Incorporated
30 January 2001 (Class 381/310); filed 10 March 1997

Head-related transfer functions (HRTFs) can be used to generate convincing phantom sound sources at any location relative to the listener. Unfortunately, such transfer functions, like fingerprints, are unique to each individual. The invention utilizes an interactive method—a brief learning session—to select the best HRTF for a particular user.—GLA

6,191,515

43.40.Tm COMBINED PASSIVE MAGNETIC BEARING ELEMENT AND VIBRATION DAMPER

Richard F. Post, assignor to The Regents of the University of California
20 February 2001 (Class 310/90.5); filed 3 September 1999

This patent pertains to a bearing system that employs arrangements of permanent magnets which act together to support a rotating element in a state of dynamic equliibrium. A piece of highly conductive nonmagnetic material, affixed to the nonrotating part of the bearing, serves to damp transverse vibrations of the rotor.—EEU

6,152,259

43.50.Gf SOUND DEADENING LAMINATE

Christoph Freist and Josef Polak, assignors to CWW-GERKO Akustik GmbH & Company, KG
28 November 2000 (Class 181/290); filed in Germany 30 January 1998

This laminate would be used for suppressing the vibration of metal surfaces, for example, car floors, dishwashers, etc. It has a bituminous sound-deadening layer more than 1 mm thick and an additional layer of polystyrene foam bonded to it for thermal insulation.—CJR

6,170,605

43.50.Gf PANEL FOR ABSORBING ACOUSTIC ENERGY

Dattatray Rao *et al.*, assignors to The Soundcoat Company, Incorporated
9 January 2001 (Class 181/290); filed 22 April 1999

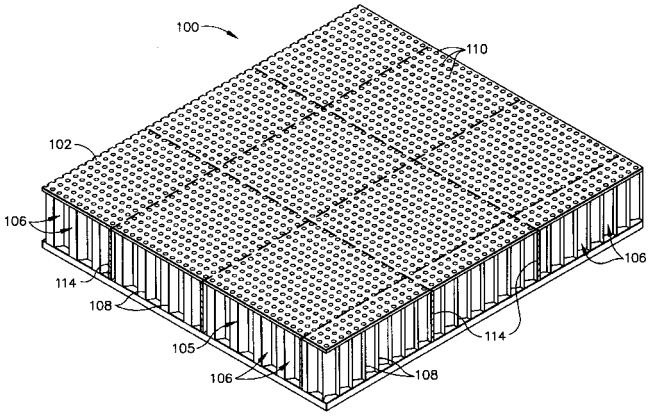
This lightweight panel is designed to absorb acoustic energy during launch of a rocket or within an engine. The panel protects the absorbing material from particulates, and a venting system allows release of gases.—CJR

6,182,787

43.50.Gf RIGID SANDWICH PANEL ACOUSTIC TREATMENT

Robert E. Kraft and Asif A. Syed, assignors to General Electric Company
6 February 2001 (Class 181/292); filed 12 January 1999

This is an acoustic treatment for the air ducts of a gas turbine engine. The treatment includes a perforated facesheet, a backplate spaced apart from the facesheet, and interconnected cells between the two. Some of the walls



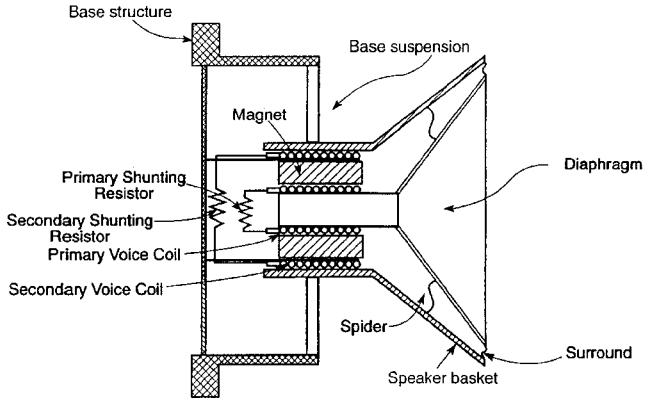
of the cells are formed of a porous material so that air is able to flow through the cells in a direction parallel to the facesheet and backplate.—CJR

6,195,442

43.50.Jh PASSIVE VIBROACOUSTIC ATTENUATOR FOR STRUCTURAL ACOUSTIC CONTROL

Steven F. Griffin and Steven A. Lane, assignors to the United States of America as represented by the Secretary of the Air Force
27 February 2001 (Class 381/396); filed 27 August 1999

This invention presents a passive vibroacoustic device that both attenuates the vibration of a flexible structure and also provides acoustic dissipation to the volume or cavity enclosed by the structure. This reduces the transmission of sound into the enclosure and reduces the vibration of the structure. By design of the shunting resistor and the mass and suspension



properties, the device can achieve high levels of both structural vibration attenuation and acoustic attenuation. Examples of acoustic cavities enclosed by a flexible structure include airplanes, trains, cars, and spacecraft launch vehicles.—CJR

6,155,116

43.55.Dt ACOUSTIC TREATMENT PERFORMANCE TESTING

John D. Muench *et al.*, assignors to the United States of America as represented by the Secretary of the Navy
5 December 2000 (Class 73/599); filed 4 October 1999

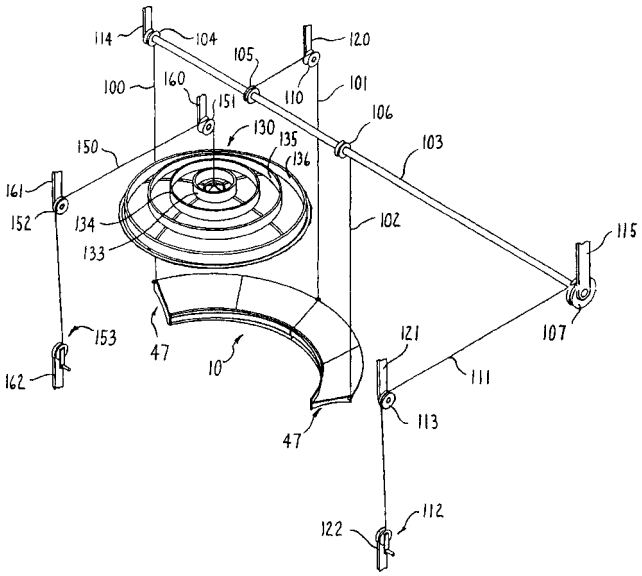
This is a procedure for field testing of acoustical performance in a free-field environment over a wide range of frequencies. The method uses a gated pulsed source aimed at the treatment being tested. The reflected energy and the gated source are compared as an indication of performance.—CJR

6,202,361

43.55.Dt OVERHEAD SUSPENDED TABLE AND ACOUSTIC DOME

Clarkson S. Thorp and Steven J. Beukema, assignors to Haworth, Incorporated
20 March 2001 (Class 52/39); filed 5 June 1998

This dome, to be suspended over a table, has a reflecting lower (concave) side to reflect sound to the work area, and an absorptive upper (outer,



convex) surface to absorb sound. The dome can be raised above or lowered into the table.—CJR

6,158,176

43.55.Ev CORE FOR A SOUND ABSORBING PANEL

Jay Perdue, Amarillo, Texas
12 December 2000 (Class 52/144); filed 6 March 1995

These rectangular acoustical panels have a core of rockwool with flat front and back faces and with a nonwoven fiberglass sheet for one of the faces. The rockwool core has a density between 5 and 9 lbs per cubic foot and a thickness between 0.75 and 2.0 in.—CJR

6,179,943

43.55.Ev METHOD FOR FORMING A COMPOSITE ACOUSTIC PANEL

John M. Welch *et al.*, assignors to The Boeing Company
30 January 2001 (Class 156/160); filed 30 July 1997

The so-called method for forming this composite panel begins by arranging a perforated sheet on a convex mandrel, laying a core material on the perforated sheet, and curing the two together. This locks in the curved shape.—CJR

6,186,270

43.55.Ev LAYERED SOUND ABSORBER FOR ABSORBING ACOUSTIC SOUND WAVES

Manfred Roller *et al.*, assignors to M. Faist GmbH & Company KG
13 February 2001 (Class 181/286); filed in Germany 14 September 1994

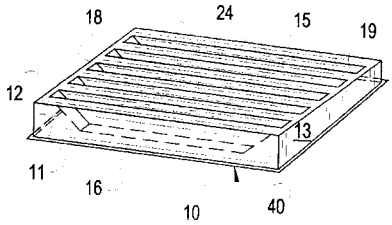
This absorber has multiple layers spaced apart from each other so that gas-filled resonance chambers are formed. The variety of different chambers, their different dimensions, various densities of the materials of the walls of the parallel chambers, thickness of the walls, and other factors give these wall panels a wide range of sound absorption frequencies.—CJR

6,209,680

43.55.Ev ACOUSTIC DIFFUSER PANELS AND WALL ASSEMBLY COMPRISED THEREOF

Jay Perdue, Amarillo, Texas
3 April 2001 (Class 181/295); filed 10 April 2000

These panels have a series of lanes or grooves that form parallel elon-



gated depressions to enhance diffusion of reflected sound in a room.—CJR

6,213,252

43.55.Ev SOUND ABSORBING SUBSTRATE

Robert Ducharme, assignor to Royal Mat International Incorporated
10 April 2001 (Class 181/293); filed 8 November 1996

This sound-absorbing substrate is made of recycled tires. It is designed to be used as an underlayment in floor systems to reduce noise transfer. The bottom surface has parallel grooves at 45 degrees with respect to the side of the panel. These grooves reduce the contact area of the substrate to the sub-floor and provide resiliency.—CJR

6,220,388

43.55.Ev ACOUSTICAL INSULATION PANEL

David M Sanborn, assignor to Strandtek International, Incorporated
24 April 2001 (Class 181/290); filed 27 January 2000

An acoustical insulation panel to absorb and attenuate sound energy consists of an inner honeycomb core and outer membranes on each side of the core. The outer membrane panels have sound-absorbing fibers that penetrate into the honeycomb core. The fabric is a nonwoven meltblown microfiber acoustical absorbing fabric.—CJR

6,170,603

43.55.Jz ACOUSTIC WALL

Wolfgang Bachmann *et al.*, assignors to Harman Audio Electronic Systems GmbH
9 January 2001 (Class 181/150); filed in Germany 4 September 1998

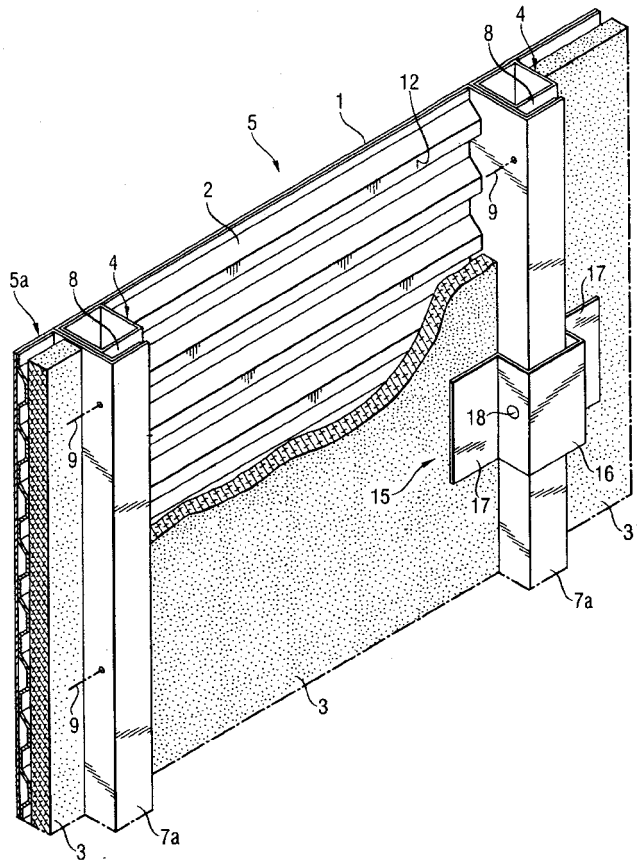
This patent describes an acoustic wall for large public events that produces sound with planar loudspeakers. Different segments of the wall allow the user to tune the response to different frequencies, and various segments of the wall can be attached to each other for control of frequency response. The electronics involved is not discussed.—CJR

6,155,016

43.55.Rg WALL STRUCTURE

Jarmo Wacker and Kari Ruusunen, assignors to Kvaerner Masa-Yards Oy
5 December 2000 (Class 52/588.1); filed in Finland 13 May 1997

This wall is designed for fire resistance and sound insulation, particularly for ships. The wall includes a surface plate that is attached to vertical



stiffeners and a thin corrugated plate that is attached to the surface plate. The result is supposed to be an effective lightweight wall.—CJR

6,196,352

43.55.Rg SOUND ATTENUATING RELOCATABLE PARTITION WALL PANEL

Jim Thompson Goodchild, assignor to SMED International Incorporated
6 March 2001 (Class 181/290); filed 14 June 1999

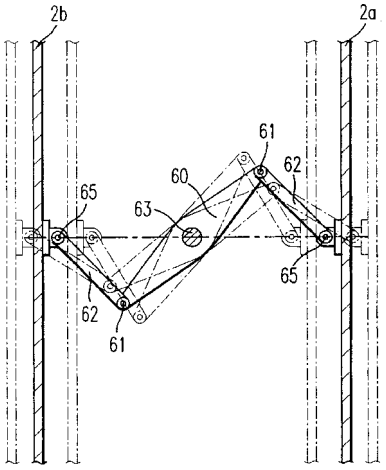
This wall design uses panels that can be attached to a frame. The panels, of different thicknesses, would reduce resonant frequencies (coincidence dip) of the assembly.—CJR

6,173,806

43.55.Ti MUFFLING WALL

Sadakuni Ito, assignor to Itoon
16 January 2001 (Class 181/210); filed in Japan 9 October 1996

This wall has a frame body and two separate diaphragms as faces. Within the wall is a mechanical linkage that reverses the phase vibration of



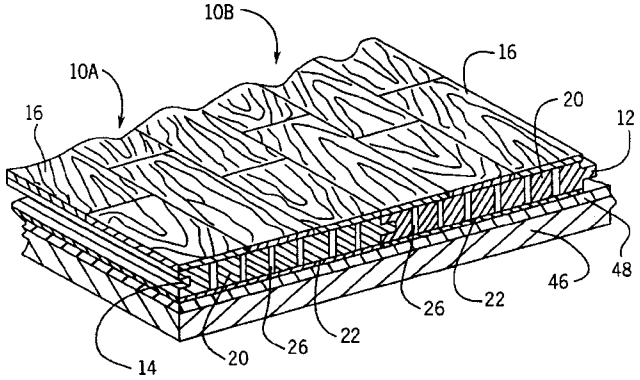
the diaphragms, thus causing the one diaphragm to be out of phase with the other, and hence diminishing the original sound.—CJR

6,182,413

43.55.Ti ENGINEERED HARDWOOD FLOORING SYSTEM HAVING ACOUSTIC ATTENUATION CHARACTERISTICS

Tryggvi Magnusson, assignor to Award Hardwood Floors, L.L.P.
6 February 2001 (Class 52/589.1); filed 27 July 1999

This floor has tongue and groove core planks. Thin strips of hardwood flooring are glued to the top of the core. The core plank itself has holes extending completely through from top to bottom. The purpose of the holes



in the core plank is to attenuate acoustic energy propagating within the core plank, and thereby reduce the level of noise heard below that is caused when people walk on the floor above.—CJR

6,183,837

43.55.Ti SOUNDPROOF ALUMINUM HONEYCOMB-FOAM PANEL

Tae Bong Kim, Jamwon-dong, Republic of Korea
6 February 2001 (Class 428/118); filed in Republic of Korea 2 June 1998

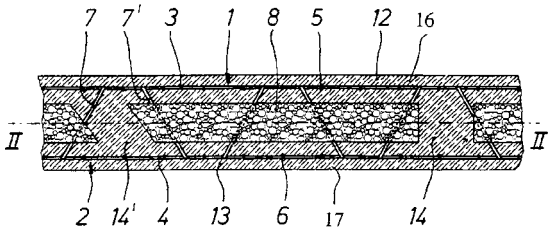
This panel can both absorb and block sound over a wide frequency range. The panel is highly durable, resistant to corrosion, and has excellent surface strength to resist impact. The base material for the panel is an aluminum foam board with open pores and a noise-insulating metal board on the back; in between is an aluminum honeycomb structure.—CJR

6,185,890

43.55.Wk BUILDING ELEMENT

Klaus Ritter, assignor to EVG Entwicklungs-u. Verwertungs-Gesellschaft m.b.H.
13 February 2001 (Class 52/309.11); filed in Austria 21 November 1996

This concrete core structure is designed to improve sound damping in prefabricated building elements having two parallel welded wire mesh mats, connected by web wires. A foam plastic insulation is fitted between the outer



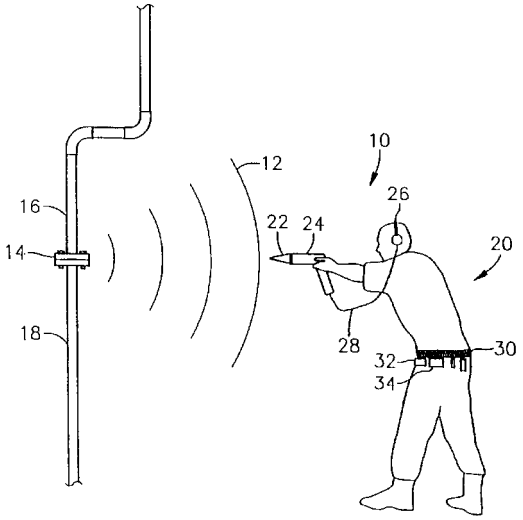
layers. Concrete is then sprayed on both outer sides over the wire mesh. This results in strong concrete shells but the system of connections keeps the system damped.—CJR

6,220,098

43.58.−e MULTIPLE SENSOR ULTRASONIC MONITORING DEVICE

William S. Johnson *et al.*, assignors to CSI Technology, Incorporated
24 April 2001 (Class 73/592); filed 6 May 1998

A holder is described to which multiple interchangable ultrasonic measuring heads may be attached. Depending on the measuring head



attached, the device configures itself to provide the appropriate measurement.—IMH

6,195,438

43.58.Kr METHOD AND APPARATUS FOR LEVELING AND EQUALIZING THE AUDIO OUTPUT OF AN AUDIO OR AUDIO-VISUAL SYSTEM

Hideki Yumoto and Theodore S. Rzeszewski, assignors to Matsushita Electric Corporation of America
27 February 2001 (Class 381/103); filed 9 January 1995

The purpose of this device is to provide automatic control of multiple-channel audio output levels. Incoming audio signals from multiple channels are combined and fed into a pair of detector-rectifier circuits. One detector-

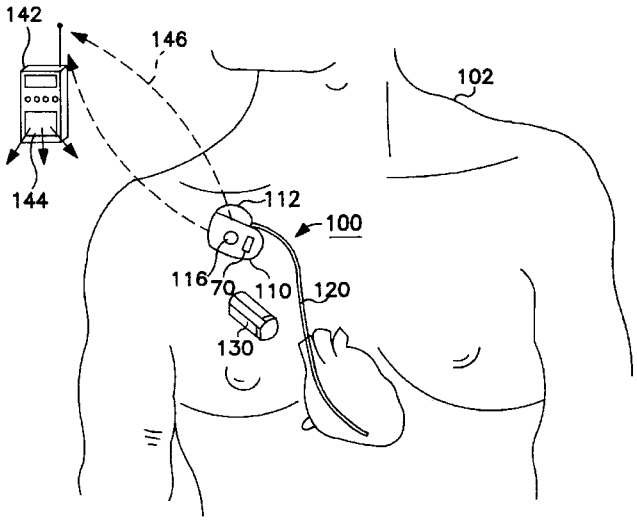
rectifier circuit acts as a smoothing circuit with a time constant longer than that of the other circuit. Signals with longer time constants (apparently those of lower frequencies) are fed to a microprocessor to control the volume settings of an audio or audio-visual system. Signals having shorter time constants (higher-frequency content) undergo source equalization.—DRR

6,216,038

43.60.Dh BROADCAST AUDIBLE SOUND COMMUNICATION OF PROGRAMMING CHANGE IN AN IMPLANTABLE MEDICAL DEVICE

Jerome T. Hartlaub *et al.*, assignors to Medtronic, Incorporated
10 April 2001 (Class 607/31); filed 29 April 1998

This implantable medical device (IMD) includes an integrated audio feedback device. The patent covers methods of gathering information from IMDs, including interrogation of programmed values, operating modes and operational conditions, confirmation of programmed changes, access to data stored in the IMD, and patient warnings. Audible sounds, including voiced



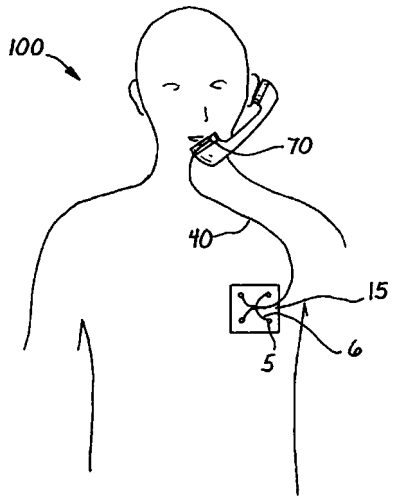
statements or musical tones stored in the analog memory, modulate a RF carrier. The broadcast radio signal is received and the audible sounds are demodulated and reproduced by a radio receiver as voiced statements or musical tones that convey warnings or status messages to the patient or another person.—DRR

6,219,408

43.60.Dh APPARATUS AND METHOD FOR SIMULTANEOUSLY TRANSMITTING BIOMEDICAL DATA AND HUMAN VOICE OVER CONVENTIONAL TELEPHONE LINES

Paul Kurth, Rancho Palos Verdes, California
17 April 2001 (Class 379/106.02); filed 28 May 1999

A system is described for simultaneous delivery of voice and biomedical data over a telephone line. In addition to carrying ordinary voice communication over the line, the system carries biomedical data which is con-



verted from sensor output into acoustic signals for transmission over the same line.—IMH

6,216,816

43.60.Qv STETHOSCOPE HAVING TWO INPUT DEVICES

T. Anthony Don Michael, Bakersfield, California
17 April 2001 (Class 181/131); filed 16 December 1999

This is a stethoscope that is conventional in most respects except that instead of only a single diaphragm, the chest-piece contains two pickups, one a metal bell and the other a diaphragm. Multiple pickups can be provided. It is said that a bell chest-piece with two pickups having specified form and relative locations is well suited for detecting heart sounds because it fits between two adjacent ribs.—DRR

6,216,025

43.60.Rw THERMOACOUSTIC COMPUTED TOMOGRAPHY SCANNER

Robert A. Kruger, assignor to Optosonics, Incorporated
10 April 2001 (Class 600/407); filed 2 February 1999

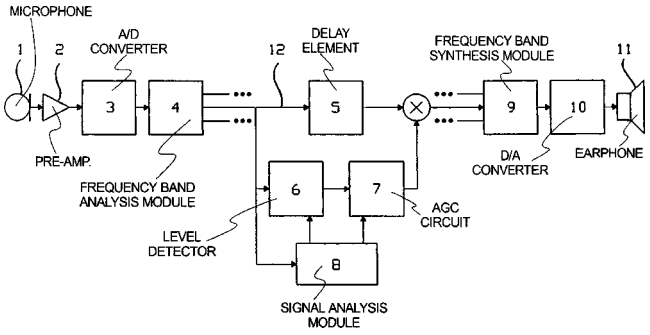
The patent describes a system for characterizing the localized electromagnetic wave absorption properties of biological tissues *in vivo* through the use of incident electromagnetic waves and analysis of the resulting acoustic waves. A number of acoustic transducers arranged on a rotatable imaging bowl are acoustically coupled to the surface of the tissue for measuring the acoustic waves which arise when the tissue is exposed to a pulse of electromagnetic radiation. The multiple transducer signals are then combined to yield an image of the absorptivity of the tissue as a possible means of medical diagnosis. Specific mathematical reconstruction procedures are described for producing images from the transducer signals. Specific arrays of transducers are described. A discussion of noise reduction techniques is also included.—DRR

6,198,830

43.66.Ts METHOD AND CIRCUIT FOR THE AMPLIFICATION OF INPUT SIGNALS OF A HEARING AID

Inga Holube *et al.*, assignors to Siemens Audiologische Technik GmbH
6 March 2001 (Class 381/321); filed in Germany 29 January 1997

Hearing aid compression parameters for achieving loudness equalization are normally determined using steady-state signals. Different listening situations are accommodated by the wearer manually switching between



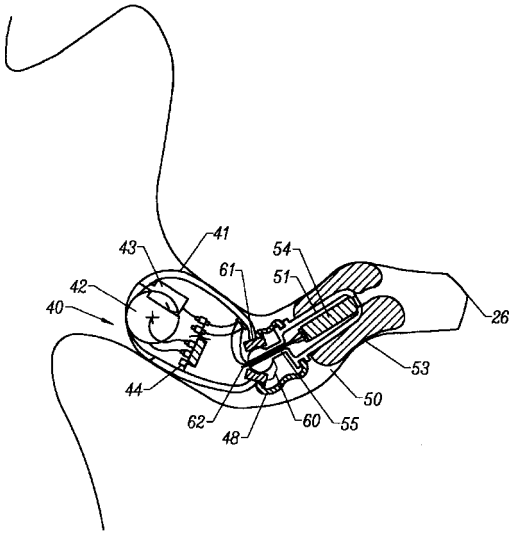
stored programs in memory. This patent discloses methodology for obtaining compression parameters using a spectral analysis of the time variation of the input signal envelope. For example, in those channels of a multi-channel hearing aid having input signals with a low modulation frequency, a high compression ratio and a long time constant may be used in the level detector. From short-term time-frequency analyses, an estimate can be made of which channels contain noise, and the signals in those channels can be attenuated.—DAP

6,212,283

43.66.Ts ARTICULATION ASSEMBLY FOR INTRACANAL HEARING DEVICES

Henry Fletcher *et al.*, assignors to Decibel Instruments, Incorporated
3 April 2001 (Class 381/313); filed 3 September 1997

Articulating a hearing aid receiver module within the ear canal is an alternative to custom-fitted earpieces for improving wearing comfort and forming an acoustic seal. This patent discloses a means for articulating a receiver module that is separated acoustically and mechanically from the



main part of a deep fitting noncustom hearing aid. A small ball-joint assembly that is almost completely contained within the main housing of the hearing aid provides independent and free movement of the receiver module and has rotational stops to limit the movement.—DAP

6,198,971

43.66.Ts IMPLANTABLE SYSTEM FOR REHABILITATION OF A HEARING DISORDER

Hans Leysieffer, assignor to IMPLEX Aktiengesellschaft Hearing Technology
6 March 2001 (Class 607/55); filed in Germany 8 April 1999

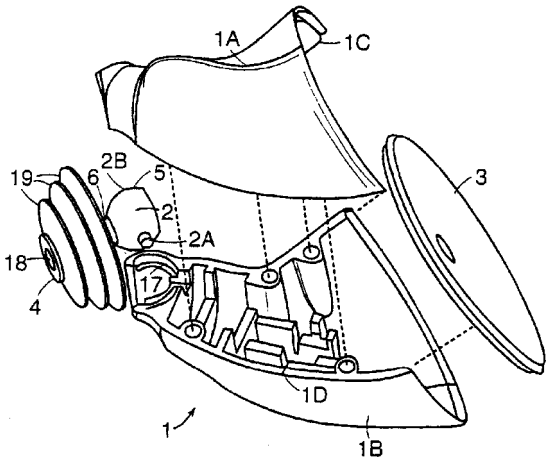
Optimally matching the performance of a cochlear or middle ear implantable system to the requirements of an individual patient may often be done only after initial implantation is completed. Thereafter, altering the operating parameters of totally implantable implants may require invasive surgery. A partially implantable system for hearing rehabilatation is described in which signal processing and signal generation parameters can be more easily adapted to the individual's performance. The invention consists of a rewritable, implantable storage means for holding an operating program that can be modified by an external unit via telemetry.—DAP

6,205,227

43.66.Ts PERITYMPANIC HEARING INSTRUMENT

Derek Dwayne Mahoney *et al.*, assignors to Sarnoff Corporation
20 March 2001 (Class 381/328); filed 31 January 1998

A deep-fitting hearing aid is described that consists of a rigid shell and a hollow, flexible tip that conforms to the ear canal. The axis of the flexible tip is movable with respect to the axis of the shell and is said to make



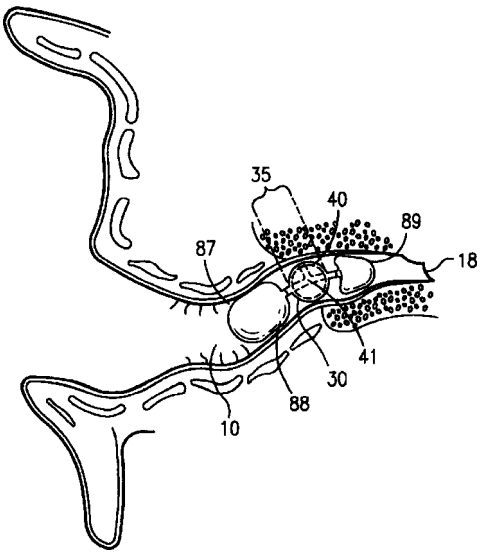
hearing aid insertion easier by following the path of the ear canal. One embodiment of the flexible tip has flanges extending radially outward that help produce an acoustic seal near the bony region of the ear canal.—DAP

6,208,741

43.66.Ts BATTERY ENCLOSURE FOR CANAL HEARING DEVICES

Adnan Shennib and Richard C. Urso, assignors to Insonus Medical, Incorporated
27 March 2001 (Class 381/323); filed 12 November 1998

Hearing aid battery compartments are normally quite bulky. A space-efficient battery enclosure and replacement method is described for application with miniature hearing devices that are placed deeply within the ear canal. The design is said to be particularly suitable for insertion and removal



of the hearing device in small and tortuous ear canals, while minimizing occlusion. The battery enclosure may be removed for disposal by stretching it over the battery or may be coated over the battery with protruding electrical contacts forming a disposable battery assembly.—DAP

6,219,427

43.66.Ts FEEDBACK CANCELLATION IMPROVEMENTS

James Mitchell Kates and John Laurence Melanson, assignors to GN ReSound as
17 April 2001 (Class 381/318); filed 18 November 1997

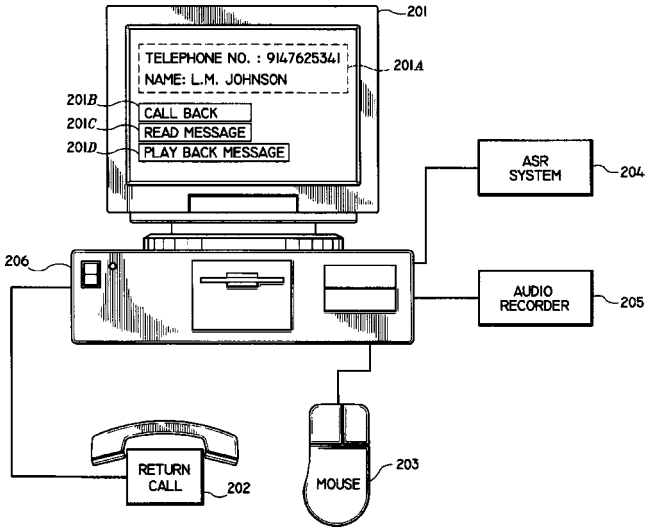
This patent describes two cascaded adaptive filters with a delay. To model the constant parts of the acoustic feedback path, the first filter uses a noise probe to adapt quickly when the hearing aid is first turned on while in the ear. The second filter does not use a probe signal but tracks perturbations in the feedback path, such as that caused by jaw movement and nearby reflective surfaces. The resulting synthesized feedback signal is subtracted from the audio signal to provide a compensated audio signal.—DAP

6,219,407

43.72.Fx APPARATUS AND METHOD FOR IMPROVED DIGIT RECOGNITION AND CALLER IDENTIFICATION IN TELEPHONE MAIL MESSAGING

Dimitri Kanevsky and Stephane Herman Maes, assignors to International Business Machines Corporation
17 April 2001 (Class 379/88.02); filed 16 January 1998

To improve automatic speaker recognition of the caller's identity in a telephone voice mail messaging system, the described text-independent speaker recognizer analyzes the digits of the caller's telephone number and the body of the message text left by the caller. The decoded information



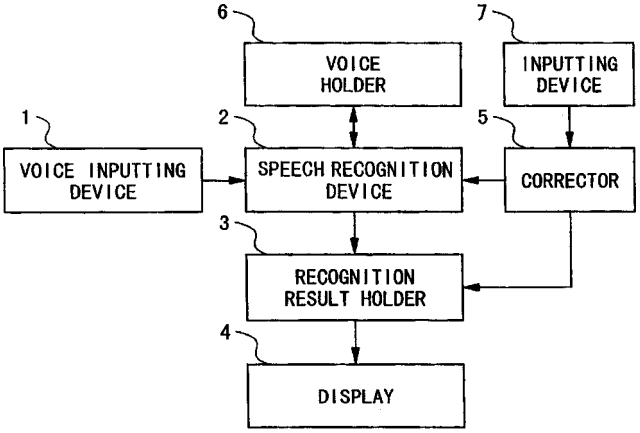
including caller's identity, associated telephone number, and the message may be presented, such as on a CRT display terminal, allowing the operator to selectively place a return call, read the message, or listen to the message.—HHN

6,212,497

43.72.Kb WORD PROCESSOR VIA VOICE

Nobumasa Araki *et al.*, assignors to NEC Corporation
3 April 2001 (Class 704/235); filed in Japan 25 November 1997

A word processor is described which recognizes words from the speech input as spoken by a user and displays the result on display **4**. Speech recognition device **2** recognizes a word by comparison with the voice data



stored in memory **6**. To increase the recognition rate, correction device **5** has access to the word processor instruction set and corrects the selected word or portion of a word.—HHN

6,195,417

43.72.Ne AUTOMATED SYSTEM FOR ACCESSING SPEECH-BASED INFORMATION

Ronald F. Dans, assignor to Telecheck International, Incorporated
27 February 2001 (Class 379/67.1); filed 18 November 1997

It had to happen. This is an automated telephone system designed to call automated telephone systems. Two primary applications are cited, the retrieval of bank account information and accessing telephone company data files to update lists of disconnected phone numbers. The system includes a speech recognizer able to understand and respond to prompts and can gen-

erate the required DTMF (Touch-Tone) codes in response. Actually, there is nothing particularly clever about the system. It requires *a priori* information about the structure of the navigable pathways for the system being called.—DLR

6,195,634

43.72.Ne SELECTION OF DECOYS FOR NON-VOCABULARY UTTERANCES REJECTION

Martin Dudemaine and Claude Pelletier, assignors to Nortel Networks Corporation
27 February 2001 (Class 704/231); filed 24 December 1997

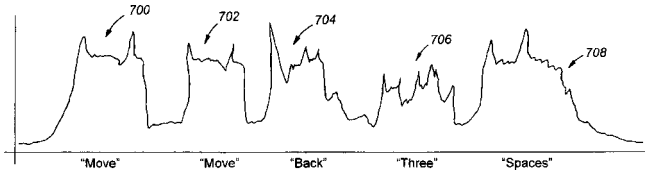
The term ''decoy,'' as used here, refers to a recognition reference pattern of a ''nonvocabulary'' nature, added to the recognition vocabulary, essentially extending the recognizable utterance set. Such patterns are also sometimes known as garbage utterances. Although the range of nonvocabulary items can be quite large, even a few such out-of-vocabulary patterns can still have a significant effect of stabilizing the recognition statistics.—DLR

6,195,635

43.72.Ne USER-CUED SPEECH RECOGNITION

Barton D. Wright, assignor to Dragon Systems, Incorporated
27 February 2001 (Class 704/231); filed 13 August 1998

This isolated-word speech recognition system is said to provide a convenient user interface for dealing with the occurrence of recognition errors. The argument is that user repetition of a falsely recognized word or phrase is faster and easier than manually selecting the correct item from a list.



The recognizer continually tests for similarity between sequential inputs. When similar items are found, such as **700** and **702**, both are processed together to help narrow the acoustical variation from that of any single utterance.—DLR

6,195,636

43.72.Ne SPEECH RECOGNITION OVER PACKET NETWORKS

Joseph A. Crupi *et al.*, assignors to Texas Instruments Incorporated
27 February 2001 (Class 704/231); filed 19 February 1999

This patent addresses the fact that speech encoding techniques usually differ depending on whether the speech is to be transmitted in coded form or to be recognized as words or phrases. A long list of well-known coding techniques is presented for each case. The patent barely mentions the crucial question of how the originating terminal makes the decision regarding the form of preprocessing to be needed.—DLR

6,195,637

43.72.Ne MARKING AND DEFERRING CORRECTION OF MISRECOGNITION ERRORS

Barbara E. Ballard and Kerry A. Ortega, assignors to International Business Machines Corporation
27 February 2001 (Class 704/235); filed 25 March 1998

When speech recognition errors occur while using this dictation system, the speaker uses either a pointing tool or speaks a distinctive command to mark the misrecognized word. Dictation can then proceed with minimal interruption. At the end of the dictated material, the marked items are brought up for further analysis, along with alternate choices or other prompts which serve to clarify the status of the failed item. Any of several correction strategies is then applied and the corrected item is inserted into the dictated material.—DLR

6,208,967

43.72.Ne METHOD AND APPARATUS FOR AUTOMATIC SPEECH SEGMENTATION INTO PHONEME-LIKE UNITS FOR USE IN SPEECH PROCESSING APPLICATIONS, AND BASED ON SEGMENTATION INTO BROAD PHONETIC CLASSES, SEQUENCE-CONSTRAINED VECTOR QUANTIZATION AND HIDDEN-MARKOV-MODELS

Stefan C. Pauws *et al.*, assignors to U.S. Philips Corporation
27 March 2001 (Class 704/256); filed in European Patent Office 27 February 1996

This patent covers a fully automatic and speaker-independent method of speech segmentation, which divides the speech signal into phonemelike segments. The method initially requires only minimal amounts of nonsegmented speech for training. According to the patent, the method is based on classifying the speech data into three broad phonetic classes (voiced, unvoiced, and silence) using sequence-constrained vector quantization and a typical application of hidden Markov models.—AAD

6,216,103

43.72.Ne METHOD FOR IMPLEMENTING A SPEECH RECOGNITION SYSTEM TO DETERMINE SPEECH ENDPOINTS DURING CONDITIONS WITH BACKGROUND NOISE

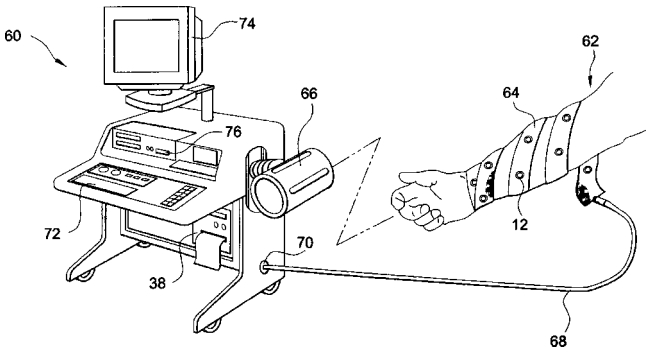
Duanpei Wu *et al.*, assignors to Sony Corporation; Sony Electronics Incorporated
10 April 2001 (Class 704/253); filed 20 October 1997

The device described in this patent determines the beginning and ending points of the noise in real-time applications. The method employs sequential short-term delta energy parameters, which are extracted by a Mel-frequency filter bank analyzer. According to the patent, all thresholds needed to decide whether a given frame of the speech signal is a begin/end point are computed from the delta energy function.—AAD

6,213,958

43.80.Ev METHOD AND APPARATUS FOR THE ACOUSTIC EMISSION MONITORING DETECTION, LOCALIZATION, AND CLASSIFICATION OF METABOLIC BONE DISEASE

Alan A. Winder, Westport, Connecticut
10 April 2001 (Class 600/586); filed 29 August 1996



The patent covers a noninvasive bone condition data acquisition system. The system measures a correlation between a wideband acoustic emission (AE) signature and a spatially localized bone microarchitecture, useful for evaluating fracture risk. The system analyzes AE signals from bone tissue and extracts time-of-arrival information and a feature vector from the analysis. These are used to obtain the locations of the AE events. The bone structure is classified using a neural network and a nearest-neighbor rule processor.—DRR

Annihilation of shocks in forced oscillations of an air column in a closed tube (L)

N. Sugimoto
Department of Mechanical Science, Graduate School of Engineering Science, University of Osaka, Toyonaka, Osaka 560-8531, Japan

M. Masuda
Department of Thermal Energy System, Ecology and Energy Systems Research Center, Sanyo Electric Company, Hashiridani, Hirakata, Osaka 573-8534, Japan

T. Hashiguchi and T. Doi
Department of Mechanical Science, Graduate School of Engineering Science, University of Osaka, Toyonaka, Osaka 560-8531, Japan

(Received 9 February 2001; accepted for publication 6 August 2001)

Effects of a periodic array of Helmholtz resonators on forced longitudinal oscillations of an air column in a closed tube are examined experimentally. The column is driven sinusoidally at a frequency near the lowest resonance frequency by oscillating bellows mounted on one end of the tube. Frequency responses are obtained for small and large amplitudes of the excitations. While the array lowers the resonance frequency and the peak value, its dispersive effect, i.e., the dependence of the sound speed on a frequency, can annihilate the shock effectively. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1407265]

PACS numbers: 43.25.+y [MFH]

A periodic array of Helmholtz resonators is very effective to annihilate a shock wave (called simply a shock hereafter) in propagation of nonlinear acoustic waves in an air-filled tube.[1–3] This is made possible by the action of wave dispersion that the array yields, whereby the sound speed in the tube tends to change depending on a frequency.[4] This letter examines whether or not the array can also annihilate a shock in forced longitudinal oscillations of an air column confined in a tube. While the theory is currently being developed, the experimental findings are reported.

Forced oscillations of the air column in a tube of uniform cross-section have been studied for several decades by many authors. An extensive review is given by Ilgamov *et al.*[5] The column is commonly driven sinusoidally by a plane piston reciprocating at a frequency close to or equal to the resonance frequency of the fundamental mode. While the amplitude of displacement of the piston is small, the linear standing wave is excited with the pressure node in the center of the tube and the loops at both ends. But as the amplitude becomes large, there emerges a shock propagating back and forth, reflected from the closed end and the piston.

When the shock appears, it suppresses the maximum excess pressure by so-called acoustic saturation and also gives rise, not only to loud noise and vibrations of the tube, but also to eventual heating of the air. To generate high-amplitude oscillations, it is therefore required to annihilate the shock. Recently new researches have begun. Lawrenson *et al.*[6] and Ilinskii *et al.*[7] have developed a novel method to achieve this by varying a tube's cross-section deliberately along the axial direction and vibrating the whole tube on the shaker. Although the driving method is different from the conventional one, their idea lies in avoidance of coincidence of the frequencies of higher harmonics of the excitation with the resonance frequencies of higher modes in the tube. When they coincide as in a uniform tube, called *consonant*,[6] the energy of the fundamental mode is easily pumped up into higher modes so that the shock is formed. By changing the cross-section, the tube is made *dissonant* and the shock may be avoided. It is emphasized that even in this dissonant tube, the system remains hyperbolic and the propagation speed given by the characteristics is not different from the sound speed. On the contrary, the connection of the array of resonators to the tube changes the hyperbolic system to a dispersive one where the phase (sound) speed depends on frequency. Thanks to this dispersion, the tube of uniform cross-section is rendered *dissonant* in consequence. But remark the difference of both mechanisms in annihilation of shocks.

The experimental setup is illustrated in Fig. 1. We employ a tube of inner diameter 80 mm, of thickness 7.5 mm, and of length 3200 mm. For the sake of comparison, we prepare another tube of the same size but without the array. Each resonator has a cavity of volume $V(=4.97\times10^{-5}\,\mathrm{m}^3)$ and a throat of length $L(=35.6\,\mathrm{mm})$ and of diameter $2r$ $(=7.11\,\mathrm{mm})$. Its lossless natural frequency $\omega_0/2\pi$ is given by $\sqrt{\pi r^2 a_0^2/L_e V}/2\pi$, where a_0 is the sound speed and L_e $(=L+2\times0.82r)$ is the throat's length with the end corrections. The resonators, 64 in total, are connected to the tube with axial spacing $d(=50\,\mathrm{mm})$ and are staggered on both sides of the tube. The parameter $\kappa(=V/Ad)$ measuring the size of the array takes the value 0.198 where A denotes the cross-sectional area of the tube.

With one end of the tube closed by a flat plate, the air column under the atmospheric pressure p_0 at room temperature is driven by bellows mounted on the other end. "By a bellows" is meant such a cylindrical tube with a flat bottom that its lateral surface is folded into conical surfaces with

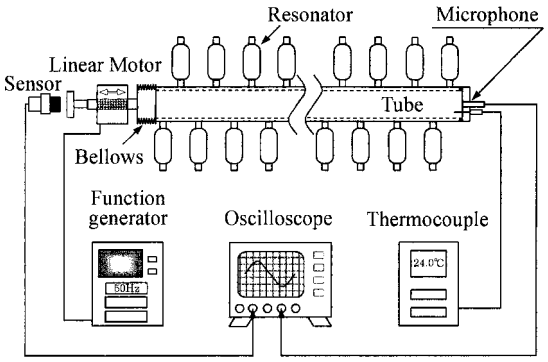


FIG. 1. Illustration of the experimental setup.

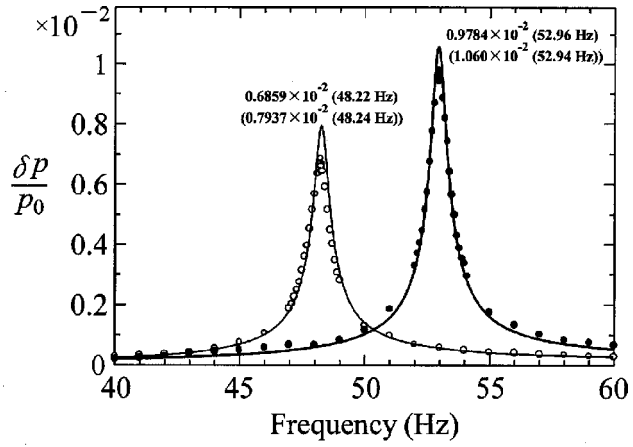


FIG. 2. Frequency responses of the fundamental mode in the tubes without the array of Helmholtz resonators and with it, respectively, for the amplitude of the displacement of the bellows 0.1 mm where the solid and blank circles indicate the experimental data measured, respectively, in the tubes without the array and with it, and the bold and thin curves represent, respectively, the amplitudes of the excess pressure Eq. (4) at the closed end; the peak values and the resonance frequencies are indicated with the theoretical values enclosed by the parentheses.

crests and troughs. Many folds of the lateral surface allow the bottom plate to move in the axial direction. The bellows is 56 mm deep in the natural state while the greatest and smallest inner diameter D_g and D_s are 110 mm and 80 mm, respectively. The bottom plate is driven at its center by a reciprocating linear motor located outside. With the depth of the bellows inclusive, the total length of the air column, l, is 3256 mm under atmospheric pressure. The displacement of the bottom plate of the bellows is measured optically by a sensor using a laser beam. Excess pressure is measured at the closed end by a microphone set flush with the flat plate at its center. The mean temperature in the tube near the closed end, T_0, is always monitored by a thermocouple.

The bellows has a merit of securing hermetic sealing whereas they have the demerit of nonuniformity in the inner diameter. To regard the displacement of the bottom plate of the bellows x_b (called simply the displacement of the bellows hereafter) as the one of the piston x_p, we make an assumption of an ''equivalent volume displacement.'' Since a wavelength of oscillations is much longer than the depth of the bellows, the axial variations of air density in the bellows may be neglected. Then, the displacement of the bellows x_b is related to that of the piston x_p by $(1+\chi+\chi^2)x_b/3$, $\chi = D_g/D_s$.[8] For $\chi = 110/80 = 1.375$, it follows that $x_p = 1.422x_b$.

We begin by describing the results when the amplitude of displacement of the bellows is small. In the lossless linear theory, the air column resonates at such an angular frequency of the driver ω that a half-wavelength is equal to the column's length. This gives the resonance frequency of the fundamental mode. Near resonance, however, small lossy effects come into play to shift the resonance frequency downward. Figure 2 shows the frequency response of the air column for the amplitude of displacement of the bellows fixed at 0.1 mm where the half of the peak-to-peak pressure, δp, is drawn relative to p_0 versus the driving frequency $\omega/2\pi$. Even at this level of excitation, the second harmonic appears but its magnitude is a few percent of the fundamental one at most. The solid circles indicate the data measured in the tube without the array under $T_0 = 24.9\,^\circ\mathrm{C}$ and $p_0 = 1.007\times10^5$ Pa. The blank circles indicate the data in the tube with the array under $T_0 = 25.6\,^\circ\mathrm{C}$ and $p_0 = 1.003\times10^5$ Pa. In the absence of the array, the frequency response has the peak value of 0.9784×10^{-2} at the resonance frequency 52.96 Hz. When the array is connected, the peak value and the frequency are lowered to 0.6859×10^{-2} and 48.22 Hz. Both the decrease in the peak value and the frequency result from the *lossless* effect of the array. The extra loss at the throats decreases the peak value. Its fraction is estimated to be merely 5.5%.

The frequency response of the second and third modes in the tube with the array is shown in Fig. 3 where the maximum acceleration, i.e., the amplitude of the displacement of the bellows times the frequency squared, is held to be a constant value at the amplitude 0.1 mm and 48.2 Hz. The blank circles indicate the experimental data. The second and third modes have, respectively, the peak values 0.2034×10^{-2} at 95.44 Hz and $T_0 = 25.4\,^\circ\mathrm{C}$, and 0.070×10^{-2} at 139.5 Hz and 25.3 °C under $p_0 = 0.999\times10^5$ Pa. The ratio of the second resonance frequency to the fundamental one 48.2

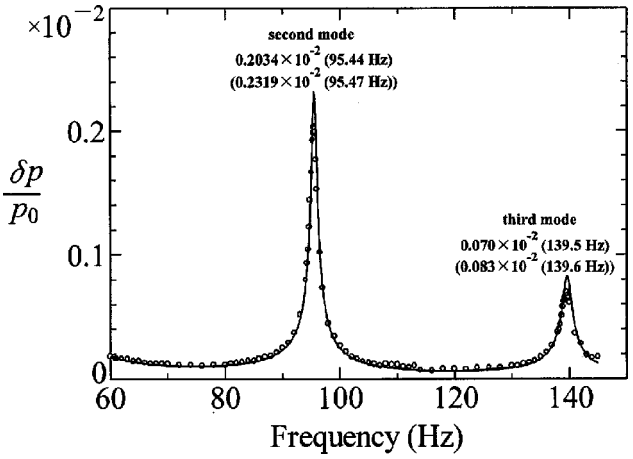


FIG. 3. Frequency response of the second and third modes in the tube with the array of Helmholtz resonators with the maximum acceleration of the bellows fixed to be the value at the amplitude 0.1 mm and 48.2 Hz where the circles represent the experimental data and the line indicates the amplitude of the excess pressure Eq. (4) at the closed end; the peak values and the resonance frequencies are indicated with the theoretical values enclosed by the parentheses.

Hz is 1.98 and very close to 2, while the ratio of the third is 2.90.

The above experimental results are explained by the linear theory, taking account of a boundary layer on the side wall of the tube. But the loss due to the diffusivity of sound is negligibly small. The oscillations are modeled by one-dimensional motions in the tube except for the boundary layer. Under the continuum approximation for the array, the excess pressure $p'(x,t)$ is governed by the following wave equation in the axial coordinate x and the time t [see Eq. (60) in Ref. 4]:

$$\frac{\partial^2 p'}{\partial t^2} - a_0^2 \frac{\partial^2 p'}{\partial x^2} + \frac{2Ca_0^2\sqrt{\nu}}{R^*} \frac{\partial^{-1/2}}{\partial t^{-1/2}} \left(\frac{\partial^2 p'}{\partial x^2} \right) = -\kappa \frac{\partial^2 p_c'}{\partial t^2}, \tag{1}$$

which is coupled with the equation for the excess pressure $p_c'(x,t)$ in the cavity [see Eq. (10) in Ref. 4]:

$$\frac{\partial^2 p_c'}{\partial t^2} + \frac{2\sqrt{\nu}}{r^*} \frac{\partial^{3/2} p_c'}{\partial t^{3/2}} + \omega_0^2 p_c' = \omega_0^2 p', \tag{2}$$

where $C = 1 + (\gamma - 1)/\sqrt{\mathrm{Pr}}$ and $c_L = (L + 2r)/(L + 2 \times 0.82r)$; γ and Pr denote, respectively, the ratio of specific heats and the Prandtl number; ν is the kinematic viscosity of air; R^* and r^* are the reduced radii of the tube and the throat, defined, respectively, by $R/(1 - r^2/2Rd)$ and r/c_L, R being the tube's radius; c_L is the factor to account for the end corrections to the throat's length.[1] Here the derivative of order $-1/2$ is defined by

$$\frac{\partial^{-1/2} p'}{\partial t^{-1/2}} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{t} \frac{1}{\sqrt{t - \tau}} p'(\tau, x) d\tau, \tag{3}$$

and the derivative of order 3/2 is defined by differentiating the one of order $-1/2$ twice with respect to t.

Suppose, for simplicity, a plane piston is driven sinusoidally at an angular frequency ω. Taking the origin of x at a mean position of the piston surface, let the displacement of the piston be $x_p = X_p \exp(-i\omega t)$ with a complex amplitude X_p, the real part being taken in the complex notation. The boundary conditions are imposed as $\partial p'/\partial x = i\rho_0 \omega U_p \times \exp(-i\omega t)$ at $x = 0$ and $\partial p'/\partial x = 0$ at $x = l$, where ρ_0 is the mean air density and U_p is a complex amplitude of the velocity given by $-i\omega X_p$. Thus Eqs. (1) and (2) are solved as follows:

$$p' = i \frac{\rho_0 \omega \cos[k(x - l)]}{k \sin kl} U_p \exp(-i\omega t), \tag{4}$$

and $p_c' = p'/\mathcal{D}_r$ with $\mathcal{D}_r = 1 - \omega^2/\omega_0^2 - (1 + i)\sqrt{2\nu/\omega}(\omega/\omega_0)^2/r^*$, where a wave number k is determined by $k = (\omega/a_0)\sqrt{(1 + \kappa/\mathcal{D}_r)/\mathcal{D}_R}$, with $\mathcal{D}_R = 1 - C(1 + i)\sqrt{2\nu/\omega}/R^*$. The derivative of order $-1/2$ of $\exp(-i\omega t)$ is given by $[(1 + i)/\sqrt{2\omega}]\exp(-i\omega t)[= (-i\omega)^{-1/2} \exp(-i\omega t)]$.

In the lossless limit ($\nu \to 0$), k is reduced to $(\omega/a_0)\sqrt{1 + \kappa/(1 - \omega^2/\omega_0^2)}$. The resonance conditions $kl = m\pi$ ($m = 1,2,3,...$) for p' give a pair of frequencies $\omega_m^{\pm}$ for a fixed value of m as solutions to the following quadratic equation in ω^2: $\omega^4 - [(1 + \kappa)\omega_0^2 + \omega_m^2]\omega^2 + \omega_0^2 \omega_m^2 = 0$, where $\omega_m (= m\pi a_0/l)$ give the lossless resonance frequencies of the mth mode in the absence of the array. For a small value of κ, the pair of the frequencies are evaluated asymptotically as $\omega_m^{\pm} = \omega_l[1 \pm \kappa/2(1 - \omega_m^2/\omega_0^2) + O(\kappa^2)]$, with $\pm$ vertically ordered and $l = 0$ for the upper sign and $l = m$ for the lower one. Each $\omega_m^{\pm}$ gives the lossless resonance frequency. By connecting the array, ω_1^- becomes lower than ω_1 because $\omega_1 < \omega_0$. When the lossy effects are taken into account, ω_1^- is further shifted down slightly.

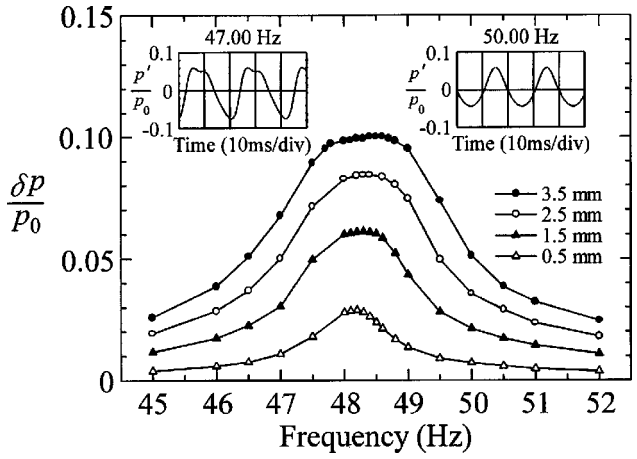


FIG. 4. Nonlinear frequency response of the fundamental mode in the tube with the array of Helmholtz resonators for the large amplitudes of displacement of the bellows $|X_b|$ indicated where the right and left insets display the temporal profiles of the excess pressure at the closed end when the column is driven at $|X_b| = 3.5$ mm and 50.00 Hz and 47.00 Hz, respectively.

To compare the experimental data against the theory, accurate values of the density ρ_0, viscosity $\mu (= \rho_0 \nu)$ and thermal conductivity k_T in $\mathrm{Pr} (= \mu c_p/k_T)$ are required, c_p being the specific heat. Using the mean temperature T_0, μ and k_T are calculated by Sutherland's formula and the one in Ref. 9 while ρ_0 is calculated from p_0 and T_0 by the equation of state for ideal gas with gas constant 2.870×10^2 J/kg K. The values of c_p and γ are taken constant as 1.007 kJ/kg K and 1.402, respectively, and a_0 is calculated by $331.5 + 0.61T_0$ m/s for T_0 measured in the degree Celsius where an effect of humidity (around 60%) is ignored.

In Fig. 2, the bold and thin curves represent the amplitude of p'/p_0 at $x = l$ calculated by Eq. (4) in the tubes without the array and with it, respectively. Here X_p is set equal to $1.422X_b$ where $|X_b|$, the amplitude of displacement of the bellows, is 0.1 mm. For the bold curve, we use $a_0 = 346.7$ m/s, $\rho_0 = 1.177$ kg/m^3, $\nu = 1.561 \times 10^{-5}$ m^2/s, and $\mathrm{Pr} = 0.7092$ at 24.9 °C while for the thin curve, $a_0 = 347.1$ m/s, $\rho_0 = 1.170$ kg/m^3, $\nu = 1.573 \times 10^{-5}$ m^2/s, and $\mathrm{Pr} = 0.7089$ at 25.6 °C so that $\omega_0/2\pi = 242.6$ Hz. The theory gives the peak value 1.060×10^{-2} at 52.94 Hz in the tube without the array, while 0.7937×10^{-2} at 48.24 Hz in the tube with the array. The experimental data agree well with the theory except for just on resonance where the maximum values measured are smaller in both cases. In Fig. 3, the curve represents the amplitude of p'/p_0 at $x = l$ with $\omega^2|X_b|$ fixed to be $(2\pi \times 48.2)^2 \times 0.1$ mm/s^2. The quantities in the parentheses are the theoretical values for the respective peaks. Good agreement between the experiment and the theory can be observed.

We now drive largely the air column in the tube with the array. Figure 4 shows the frequency response with $|X_b|$ fixed

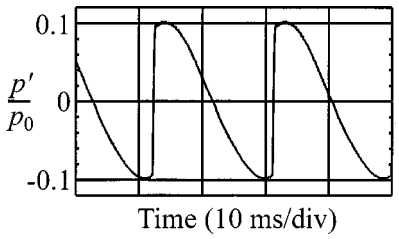


FIG. 5. Temporal shock profile in the excess pressure p' relative to p_0 at the closed end of the tube without the array of Helmholtz resonators.

at 0.5, 1.5, 2.5, and 3.5 mm, respectively, where $T_0 = 25.6\,^{\circ}\mathrm{C}$ and $p_0 = 1.007 \times 10^5$ Pa. The half of the peak-to-peak pressure, δp, is plotted relative to p_0 against $\omega/2\pi$. As the driving amplitude increases, the curves become asymmetric with a flatter peak, and the resonance frequencies become slightly higher. Incidentally, the second harmonics in the displacement of the bellows tend to appear. Its magnitude is below 4% at the greatest excitations. For $\delta p/p_0 = 0.1$, the sound pressure level attains 170 dB. But the pressure profile remains free from the shock. In the tube without the array, the shock appears even for smaller excitations. In fact, when the air column is driven at the amplitude $|X_b| = 2.423$ mm and 52.96 Hz, the shock is observed. Figure 5 shows the temporal profile of the excess pressure p', relative to p_0, measured at the closed end. In passing, this is the case just on resonance because each "N" profile is anti-symmetric with respect to the point $p'/p_0 = 0$.[10]

By contrast, Fig. 6 shows the temporal profiles of p'/p_0 at the closed end of the tube with the array when the column is driven at $|X_b| = 3.5$ mm and $\omega/2\pi = 48.50$ Hz. The profile

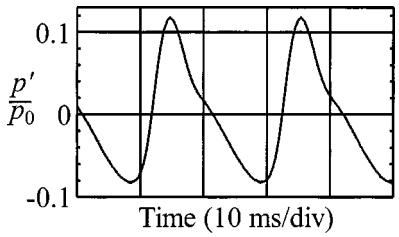


FIG. 6. Temporal shock-free profile in the excess pressure p' relative to p_0 at the closed end of the tube with the array of Helmholtz resonators.

is smooth but not symmetric with its peak. When the frequency is increased to 50.00 Hz, the profile becomes symmetric, which is shown in the right inset of Fig. 4. When the frequency is lowered to 47.00 Hz, the peaks are flattened as in the left inset. Such a flat peak appears in a very narrow range of frequencies around 47.0 Hz. Over the frequencies in between, the profiles are similar to the asymmetric one shown in Fig. 6. As the frequencies are set further away from the resonance frequency, the profiles tend to take a symmetric form but are smaller in magnitude. Hence, the array of Helmholtz resonators is also shown to be effective to annihilate the shock in the case of oscillations as well. This provides another new method to generate high-amplitude and shock-free oscillations of a gas column.

The authors wish to thank T. Nabeshima for his assistance in the experiment. This work has been supported by the Grants-in-Aid from the Japan Society for the Promotion of Science and also from The Mitsubishi Foundations, Tokyo, Japan.

[1] N. Sugimoto, "Propagation of nonlinear acoustic waves in a tunnel with an array of Helmholtz resonators," J. Fluid Mech. **244**, 55–78 (1992).
[2] N. Sugimoto, "Acoustic solitary waves in a tunnel with an array of Helmholtz resonators," J. Acoust. Soc. Am. **99**, 1971–1976 (1996).
[3] N. Sugimoto, M. Masuda, J. Ohno, and D. Motoi, "Experimental demonstration of generation and propagation of acoustic solitary waves in an air-filled tube," Phys. Rev. Lett. **83**, 4053–4056 (1999).
[4] N. Sugimoto and T. Horioka, "Dispersion characteristics of sound waves in a tunnel with an array of Helmholtz resonators," J. Acoust. Soc. Am. **97**, 1446–1459 (1995).
[5] M. A. Ilgamov, R. G. Zaripov, R. G. Galiullin, and V. B. Repin, "Nonlinear oscillations of a gas in a tube," Appl. Mech. Rev. **49**, 137–154 (1996).
[6] C. C. Lawrenson, B. Lipkens, T. S. Lucas, D. K. Perkins, and T. W. Van Doren, "Measurements of macrosonic standing waves in oscillating closed cavities," J. Acoust. Soc. Am. **104**, 623–636 (1998).
[7] Y. A. Ilinskii, B. Lipkens, T. S. Lucas, T. W. Van Doren, and E. A. Zabolotskaya, "Nonlinear standing waves in an acoustical resonator," J. Acoust. Soc. Am. **104**, 2664–2674 (1998).
[8] When a piston is displaced by x_p into the tube, it displaces volume of Ax_p. When the bottom plate of the bellows is displaced by x_b, it displaces volume of $\pi(D_s^2 + D_s D_g + D_g^2)x_b/12$, if we assume the bellows to shrink or expand uniformly. By equating both volume with $A = \pi D_s^2/4$, we derive the equivalent displacement of the bellows.
[9] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, New York, 1991), pp. 513–514.
[10] W. Chester, "Resonant oscillations in closed tubes," J. Fluid Mech. **18**, 44–64 (1964).

The effect of exit radii on intraglottal pressure distributions in the convergent glottis (L)

Ronald C. Scherer[a)]
Department of Communication Disorders, Bowling Green State University, 200 Health Center, Bowling Green, Ohio 43402

Kenneth J. De Witt
Department of Communication Disorders, Bowling Green State University, 200 Health Center, Bowling Green, Ohio 43402 and University of Toledo, Chemical Engineering, 3060 Nitschke Hall, Toledo, Ohio 43606-3390

Bogdan R. Kucinschi
Mechanical, Industrial, and Manufacturing Engineering, University of Toledo, 4006D Nitschke Hall, Toledo, Ohio 43606-3390

(Received 11 September 2000; accepted for publication 13 August 2001)

Phonation depends upon the dynamic distribution of glottal air pressures that act upon the vocal folds, as well as tissue properties of the vocal folds. The glottal wall pressures depend upon the shape, size, and diameter of the glottis for a given flow. This study examined how the radius of curvature of the glottal exit in the converging glottis affects the wall pressures in the glottis. The following exit radii were used: 0.0908 cm, 0.0454 cm, and 0.0050 cm for the 10° case, and 0.0841 cm, 0.021 025 cm, and 0.0050 cm for the 20° case. Minimal glottal diameter and flow were held constant at 0.02 cm and 73.2 cm^3/s, respectively. The computational fluid dynamics code FLUENT was used to obtain the pressure profiles. Both the transglottal and intraglottal pressures increased as the exit radius decreased, resulting in an increase in flow resistance and an increase in the outward pressure forces on the vocal folds. The results suggest that the glottal exit curvature should be well specified when building computational or physical models of the vocal folds. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1408255]

PACS numbers: 43.70.Aj [AL]

I. INTRODUCTION

Modeling laryngeal aerodynamics from a quasi-steady point of view (e.g., van den Berg *et al.*, 1957) requires specification of the glottal geometry. This includes not only the shape of the glottis (angle, symmetry, and linearity), the inferior–superior and anterior–posterior lengths of the glottal duct, and the minimal glottal diameter, but also the curvature of the vocal folds at the glottal entrance and exit locations (e.g., Scherer *et al.*, 2001). This research note examines the effects on the transglottal and intraglottal wall pressures when the exit radius of the convergent glottis is changed. The convergent glottis occurs during glottal opening during the phonatory cycle. If the (previously unstudied) variation of the exit radius creates substantial changes to the transglottal and glottal wall pressures, then the motion of the vocal folds and the shape and size of the glottal volume velocity may be significantly affected. Physical and numerical models of the glottis and phonation would need to recognize or take into consideration these effects.

II. METHODS

This study used a commercially available computational fluid dynamics code called FLUENT (Fluent, Inc., Lebanon, NH) to study the effects of constant flows through static glottal shapes. The code numerically solves the Navier–Stokes equations for the air flow physics through the specified configurations using the control-volume technique. Grids used in the present simulation contained both structured and unstructured meshes. The mesh size varied from 106k to 173k nodes and was not uniform throughout. The mesh was made finer in regions where high gradients in pressure and/or velocity occurred. All computational runs were performed unsteady which allowed a gradual transition to a steady state solution with absolute residuals of 10^{-5} for each variable (mass and velocity components). All runs were made using double precision. This code has been shown to produce intraglottal pressures for constant flows in laryngeal ducts that were highly similar to (and often negligibly different from) those found empirically using a Plexiglas model of the larynx (Scherer *et al.*, 2001). In this study, the flow field was assumed to be symmetric across the midline of the glottis, and therefore only the half flow field was modeled.

The convergent glottal angles were 10° and 20°. Three glottal exit radii were used for each angle. For the 10° case, the exit radii were 0.0908 cm, 0.0454 cm, and 0.0050 cm, and for the 20° case, 0.0841 cm, 0.021 025 cm, and 0.0050 cm. Held constant were the glottal length (1.2 cm), glottal axial flow duct length (0.30 cm), the minimal glottal diameter (0.02 cm), the glottal entrance radius (0.15 cm), the subglottal entrance angle (50° relative to the flow axis), the up-

a)Author to whom correspondence should be addressed; electronic mail: ronalds@bgnet.bgsu.edu

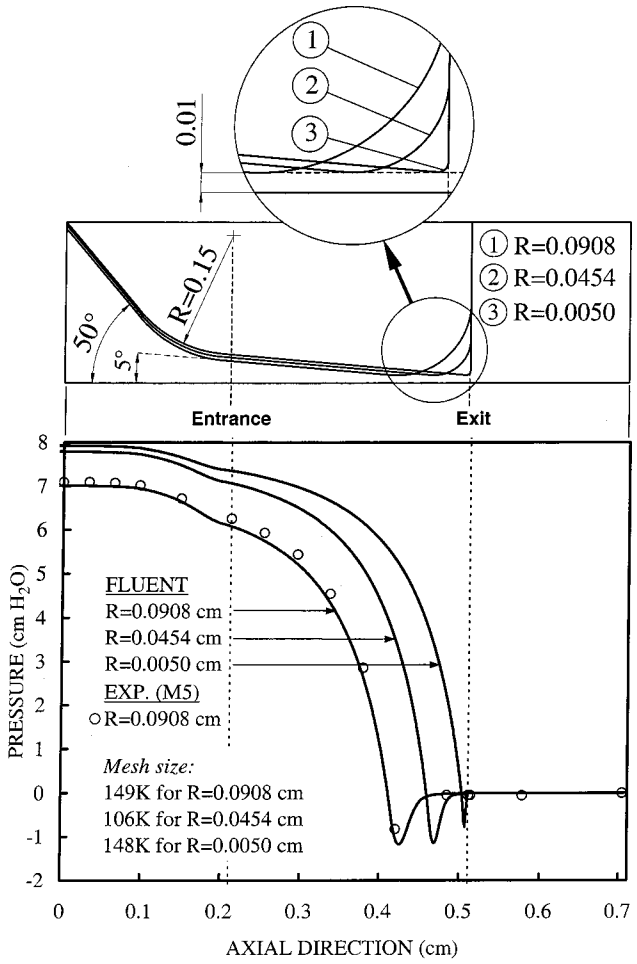


FIG. 1. Convergent 10° case (5° half-angle). The upper portion shows the outlines of the glottal configurations corresponding to the three exit radii (the entire vocal fold is not shown). Distances are in cm. The lower portion gives the pressure profiles predicted by FLUENT as well as the Plexiglas model M5 data for the largest exit radius condition. The flow used in the three FLUENT conditions was 73.2 cm^3/s. The flow for the empirical model M5 data was 65.2 cm^3/s.

stream axial glottal convergence length (0.875 cm), and the outlet domain axial length (0.2 cm). The superior surface of the vocal folds was perpendicular to the flow axis. The false vocal folds were not included. Results will be shown for a flow value of 73.2 cm^3/s, giving a Reynolds number of 800 based on the minimal glottal diameter.[1] The flow of 73.2 cm^3/s gave reasonable transglottal pressures expected during normal phonation (approximately 7 to 8 cm H_2O).

For each convergent angle case, one geometry (the one with the largest exit radius) was identical to that of the Plexiglas model M5. The results for that condition will include the FLUENT predictions and the model M5 pressure data.[2]

III. RESULTS

Figure 1 shows the results for the 10° convergence case. The upper portion of the figure gives the configurations for the glottis, including the three vocal fold radii that defined the glottal exit expansion. The pressure profiles in the lower portion of the figure show the FLUENT results. Included are pressure tap values from the Plexiglas model M5 for the largest radius; the excellent match suggests validity for the pressure profiles for the other two radius conditions as well (and similarly for the 20° case).

Both the transglottal pressure (seen as the pressure on the *y*-axis) and the pressures of the glottal wall profiles increased as the glottal exit radius became smaller. Comparing the pressures between the largest and the smallest radius conditions, the transglottal pressure rose by 13% and the pressure at glottal entrance increased by 21%. The exit curvature created a pressure profile dip due to flow acceleration in the converging section and the subsequent pressure rise within the exit expansion. The location of the minimal diameter and the pressure dip moved more downstream as the radius became smaller, thus strongly influencing the pressure profiles *upstream* of the dip. At the point where the largest radius condition had its greatest dip (at approximately 0.425 cm axial distance), the difference between the pressures for the largest and smallest radii conditions was 6.32 cm H_2O, a large difference. Indeed, if the pressure were zero (atmospheric) at the outlet to the glottis, the pressure would be negative (-1.17 cm H_2O) for the largest radius condition, but positive (5.15 cm H_2O) for the smallest radius condition at that 0.425 cm axial location, suggesting not only a significant difference in the pressures, but a difference in pulling or pushing the vocal folds (a polarity effect).

Figure 2 shows similar results for the 20° case. The sensitivity to the transglottal pressure is different from the 10° case, but suggests again that the transglottal pressure increased with smaller exit radii. The increase in transglottal pressure was 6.8% between the largest and smallest radius conditions. The pressure difference for these two conditions at the point of maximum pressure dip for the larger radius was 7.3 cm H_2O. Again, the smaller radius moved the minimal diameter and pressure dip locations more downstream, migrating the pressure profile downstream also, and creating greater positive pushing pressures on the vocal fold surface.

IV. DISCUSSION

In general, as the exit radius became smaller and the corresponding minimum diameter and minimal pressure location moved more downstream, the pressure profile expanded further downstream, creating greater outward pushing forces on the vocal folds.

The rounded glottal exit expansion, expected to be present because the normal vocal fold is soft tissue, created a pressure dip prior to the glottal exit. This dip creates a region in which the pressure on the vocal fold is less than the prevailing supraglottal pressure, enhancing the upstream–downstream pressure variation and any resulting inferior–superior glottal phasing.

The pressure drops shown by the pressure profiles result from the total drag, which is composed of friction drag and form drag. Friction drag dominates in the converging glottal section, whereas form drag constitutes the majority of the drag in the glottal exit region due to the adverse pressure gradient causing flow separation. The 20° case has a wider glottal entrance than the 10° case, and thus has less friction drag (for the same flow) and retains higher pressures in the upstream half of the glottis. For either glottal angle, the

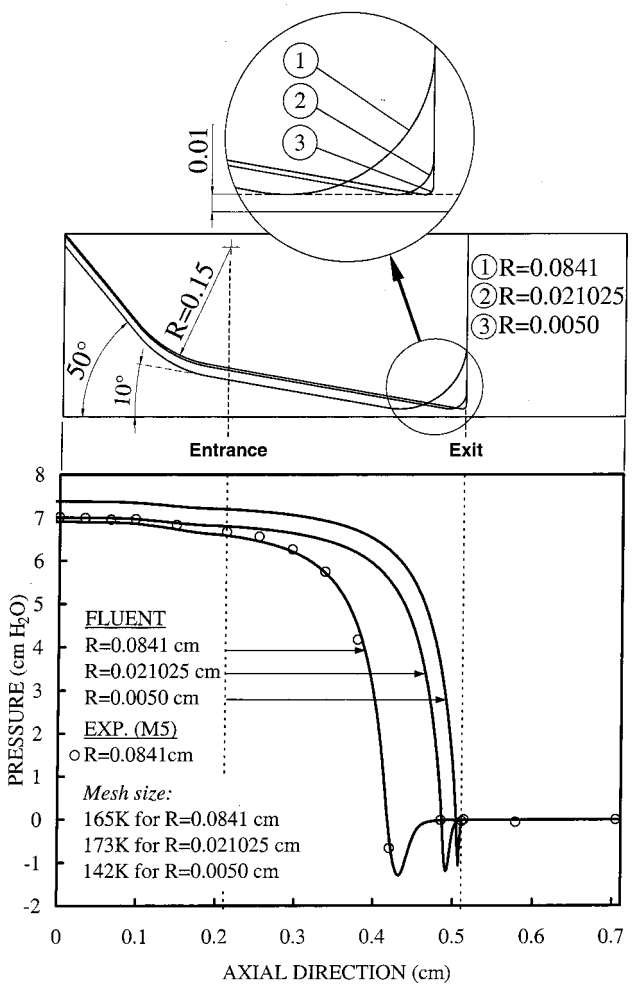


FIG. 2. Convergent 20° case (10° half-angle). The upper portion shows the outlines of the glottal configurations corresponding to the three exit radii (the entire vocal fold is not shown). Distances are in cm. The lower portion gives the pressure profiles predicted by FLUENT as well as the Plexiglas model M5 data for the largest exit radius condition. The flow used in the three FLUENT conditions was 73.2 cm^3/s. The flow for the empirical model M5 data was 68.7 cm^3/s.

transglottal pressure was reduced as the exit radius became larger because the greater rounding produced a reduction in the form drag due to the greater length over which the adverse pressure gradient occurred, and a reduction of the length over which friction drag acted.

Contrasting the $R=0.0050$ cm and $R=0.0454$ cm conditions for the 10° geometry, the decrease in the transglottal pressure due to the form drag reduction was slight (approximately 2%) because the glottal friction drag still dominated over about 75% of the glottal length until the pressure increased due to the exit radius. Conversely, for the 20° geometry and contrasting the $R=0.0050$ cm and $R=0.021\,025$ cm conditions, the transglottal pressure drop was reduced by about 5.14% by the more dominant form drag decrease.

Contrasting the $R=0.0454$ cm and $R=0.0908$ cm condition for the 10° geometry, the increased outlet radius reduced the form drag and the effective friction drag surface length such that there was a 10% decrease in the overall transglottal pressure drop. The increased exit radius caused the minimum pressure to move upstream in the glottal section, allowing a more gradual pressure increase (adverse pressure gradient) to occur in the glottal exit expansion, resulting in less form drag. This latter behavior was also evident for the 20° case.

These results suggest that models of glottal aerodynamics need to take into consideration the glottal exit curvature, especially as the exit curvature may reflect normal to diseased states of the vocal fold mucosa. For example, edema would tend to round the vocal fold exit curvature, and certain epithelial lesions may sharpen the superior edge.

V. CONCLUSIONS

During phonation, the convergent glottis occurs during glottal opening when the pressures are positive to push the vocal folds away from each other. The results of this study suggest that pressure magnitudes and the shape of the pressure distribution on the glottal walls are highly dependent upon the radius of the glottal exit curvature for the convergent glottis. A smaller radius places the minimal glottal diameter more downstream, moving the pressure profile more downstream, increasing the pressures in the glottis, and increasing the form drag effects. Therefore, the exit radius and its pressure effects should be taken into account when using physical, mathematical, and computer phonatory models that are designed to be (or inherently are) sensitive to glottal geometry. This study was meant to highlight the issue of the pressure effects due to glottal exit radii, and a more complete elucidation of the problem is necessary.

ACKNOWLEDGMENT

This report was supported by NIH Grant No. 1 R01 DC03577.

[1]The Reynolds number was defined as $\mathrm{Re}=D_hV/v$, where D_h was the hydraulic diameter, equal to 4 times the cross sectional area ($LD=1.2\,\mathrm{cm}\times0.02\,\mathrm{cm}$) divided by the perimeter ($2(L+D)$), V was the average velocity, equal to the volume flow divided by the minimal glottal cross sectional area, and v was the kinematic viscosity (0.15 cm^2/s).

[2]Model M5 (Scherer *et al.*, 2001) had been run for the angles and diameter used in this study for transglottal pressures of 3, 5, 10, and 15 cm H_2O. FLUENT was used to predict a transglottal pressure for the largest exit radius condition (the same radius as in the Plexiglas model M5) for the prescribed convergent angles and a flow of 73.2 cm^3/s. The FLUENT transglottal pressures were used then to determine the corresponding flows in model M5 that would have given rise to those transglottal pressures (the flows were found to be slightly less than in the FLUENT predictions). Those flows were then used to estimate the pressures at each pressure tap in model M5. Those pressures are plotted in the figures.

van den Berg, Jw., Zantema, J. T., and Doornenbal, P. (**1957**). "On the air resistance and the Bernoulli effect of the human larynx," J. Acoust. Soc. Am. **29**, 626–631.

Scherer, R. C., Shinwari, D., DeWitt, D., Zhang, C., Kucinschi, B., and Afjeh, A. A. (**2001**). "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," J. Acoust. Soc. Am. **109**, 1616–1630.

Evaluation of a fundamental integral in rough-surface scattering theory

David M. Drumheller[a] and Robert F. Gragg[b]
Naval Research Laboratory, Acoustics Division, Code 7140, Washington, D.C. 20375-5350

(Received 11 September 2000; revised 14 June 2001; accepted 20 August 2001)

An algorithm is presented for the numerical evaluation of a fundamental but intractable integral that occurs in the physical theory of scattering from random rough surfaces. It is based on a rational-function approximation to an integrand factor, augmented with techniques for excluding poles and zeros from the path of integration. Examples, complete with error analysis, are provided for cases relevant to acoustic sea-floor and sea-surface scattering. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1412445]

PACS numbers: 43.20.Fn, 43.30.Hw [SAC-B]

I. INTRODUCTION

In theoretical analyses of bistatic scattering from random rough interfaces, the nonspecular scattering strength per unit area is often expressed as the product of an algebraic factor and a scattering integral of the form

$$I(\alpha)=\int_0^\infty J_0(y)\, y\, B(y,\alpha)\, dy \tag{1a}$$

containing the exponential factor

$$B(y,\alpha)=\exp(-\alpha y^{2\nu}). \tag{1b}$$

This formulation pertains to both impenetrable and penetrable surfaces. In ocean acoustics it applies, for example, to small-slope scattering from a pressure-release sea surface,[1] to Kirchhoff scattering from a fluid bottom[2] and to small-slope scattering from a fluid or elastic sea floor.[3–5] Two assumptions are implicit: that the interface roughness, measured from a mean horizontal plane, is a uniform isotropic Gaussian random process; and that its spatial spectrum has the power-law form[6] $S(k)\propto k^{-\gamma}$ with $2<\gamma<4$. The parameter $\nu=\gamma/2-1$ is simply the spectral exponent γ scaled to fit within the range $0<\nu<1$, and

$$\alpha=\frac{(h_{\rm rms}Q_z)^2}{(2h_0|\boldsymbol{Q}|)^{2\nu}}\times\frac{\Gamma(1-\nu)}{\Gamma(1+\nu)}, \tag{2}$$

in which $h_{\rm rms}$ and h_0 are the rms value and correlation length of the roughness, and $\boldsymbol{Q}$ and Q_z are the horizontal and vertical components of the wave vector difference (incident minus scattered). The value of α is proportional to $f^{4-\gamma}$ and varies greatly with geometry, a typical range of interest for sea-floor and sea-surface backscattering at a few kilohertz being $10^{-3}<\alpha<10^2$ (Sec. III).

The integral $I(\alpha)$, though well defined and simply expressed, is not easily evaluated. A closed form exists only in one special case (Sec. III C) and numerical evaluation with any generality is difficult because the integrand, in addition to oscillating, decays at a highly variable rate. In practical applications, this difficulty is compounded by the additional requirement that $I(\alpha)$ be evaluated over a wide range of α values, i.e., a large spread of grazing angles. There appears to be only one approach to a solution detailed in the literature,[1,7] a technique that fuses a pair of separate series approximations for small and large α. The usefulness of this technique is limited, however, because the large-α series is restricted to its region of convergence, $\gamma>3$. For γ slightly above 3, a great many terms are required for sufficient accuracy (e.g., 6440 terms for $\gamma=3.1$), which can lead to considerable slowdown in numerical implementations.

This article presents a new method, based on a rational-function approximation, that yields $I(\alpha)$ for intermediate to large values of α (i.e., all but rather small grazing angles). The result is well suited for implementation in a vectorized computing environment such as MATLAB, provides a unified means of calculating the scattering integral throughout the γ range of practical interest in ocean acoustics (roughly $2.4\leqslant\gamma\leqslant3.9$, i.e., $0.2\leqslant\nu\leqslant0.95$) and suffers no numerical slowdown when γ is near 3. The algorithm is developed in Sec. II, examples of its use are presented in Sec. III and the summary and conclusions follow in Sec. IV.

II. ALGORITHM DEVELOPMENT

A. $I(\alpha)$ from a single rational-function approximation

Suppose, for the present, that a reliable rational-function approximation to $B(y,\alpha)$ is available for the parameter value $\alpha=1$. In terms of the variable $x=y^2$, this is $P(x)\approx B(y,1)$ where $P(x)$ is a ratio of two polynomials,

$$P(x)=\underbrace{\frac{1+\Sigma_{\ell=1}^{L}a_\ell x^\ell}{1+\Sigma_{m=1}^{M}b_m x^m}}_{\text{rational polynomial form}}=\underbrace{\frac{\Pi_{\ell=1}^{L}(1-x/z_\ell)}{\Pi_{m=1}^{M}(1-x/p_m)}}_{\text{pole-zero form}}=\underbrace{\sum_{m=1}^{M}\frac{r_m}{x-p_m}}_{\text{partial fraction form}}, \tag{3}$$

with $M-L>0$ so that $P(x)\to0$ as $x\to\infty$. The first form involves the polynomial coefficients, a_ℓ and b_m; the second involves their roots, the zeros z_ℓ and poles p_m; the third, valid when the poles are all simple,[8] involves the poles and the residues r_m.

[a]Electronic mail: David.Drumheller@nrl.navy.mil
[b]Electronic mail: Robert.Gragg@nrl.navy.mil

The key to extending this to $\alpha \neq 1$ is the observation that, owing to the form of Eq. (1b), $B(y,\alpha)$ obeys the scaling law $B(y,\alpha)=B(\alpha^{1/2\nu}y,1)$, which allows the partial fraction form of Eq. (3) to be converted into an expression for general α,

$$B(y,\alpha)\approx\sum_{m=1}^{M}\frac{\alpha^{-1/\nu}r_m}{y^2-\alpha^{-1/\nu}p_m}. \tag{4}$$

With Eq. (4) and the standard integral $\int_0^\infty yJ_0(y)/(y^2+q^2)\,dy=K_0(q)$, which is valid for $\mathrm{Re}(q)>0$,[9] Eq. (1a) reduces to

$$I(\alpha)\approx\sum_{m=1}^{M}\alpha^{-1/\nu}r_mK_0(\sqrt{-\alpha^{-1/\nu}p_m}) \quad \text{for } \mathrm{Re}(\sqrt{-p_m})>0\,. \tag{5}$$

Although this appears to afford a fairly direct means of evaluating the scattering integral, some details of the rational-function approximation require further attention. First, $P(x)$ can only approximate $B(\sqrt{x},1)$ reliably on a finite interval, and that interval has not been specified yet. Second, if the approximation is to be achieved by matching $P(x)$ to $B(\sqrt{x},1)$ at points within that interval, these points may need to be chosen carefully because, as will be demonstrated shortly, there can be only $L+M$ of them. Third, once a prospective $P(x)$ is obtained, some way must be found to insure that all of its poles meet the condition $\mathrm{Re}(\sqrt{-p_m})>0$. Fourth, since $B(\sqrt{x},1)$ is positive definite for $x>0$, the zeros of $P(x)$ should never be allowed to fall on the positive x axis. And, finally, suitable polynomial orders, L and M need to be determined. These issues are addressed in Secs. B–F below.

B. Approximation of $B(\sqrt{x},1)$

$P(x)\approx B(\sqrt{x},1)$ must be accurate throughout some finite *approximation interval*, $0\leqslant x\leqslant x_{\max}$. This is achieved by requiring that $P(x)=B(\sqrt{x},1)$ at the origin and at $N=L+M$ other *interpolation points* x_i within the approximation interval. This leads to the following N linear equations:

$$e^{-x_i^\nu}\left(1+\sum_{m=1}^{M}b_mx_i^m\right)=1+\sum_{\ell=1}^{L}a_\ell x_i^\ell \quad \text{for } i=1,\ldots,N. \tag{6}$$

Thus in matrix notation,

$$\begin{bmatrix} x_1 & \cdots & x_1^L & -x_1e^{-x_1^\nu} & \cdots & -x_1^Me^{-x_1^\nu} \\ \vdots & & \vdots & \vdots & & \vdots \\ x_N & \cdots & x_N^L & -x_Ne^{-x_N^\nu} & \cdots & -x_N^Me^{-x_N^\nu} \end{bmatrix}\begin{bmatrix} a_1 \\ \vdots \\ a_L \\ b_1 \\ \vdots \\ b_M \end{bmatrix} = \begin{bmatrix} e^{x_1^\nu}-1 \\ \vdots \\ e^{x_N^\nu}-1 \end{bmatrix} \tag{7}$$

must be solved for the N polynomial coefficients a_ℓ and b_m. Once these are known, a polynomial root-finding algorithm can calculate the poles and zeros of $P(x)$. These, in turn, yield the residues through the standard expression[10]

$$r_m=-p_m\frac{\prod_{j=1}^{L}(1-p_m/z_j)}{\prod_{\substack{\ell=1\\ \ell\neq m}}^{M}(1-p_m/p_\ell)} \quad \text{for } m=1,\ldots,M. \tag{8}$$

Solutions to Eq. (7), however, do not always yield suitable rational-function approximations. Care must be used in choosing the approximation interval and in distributing the interpolation points within it. Furthermore, the poles and zeros obtained in this way may have to be modified to yield stable, accurate approximations to B and I. Procedures are presented in Secs. C–F that reliably accomplish these objectives in practice.

C. Approximation interval and interpolation points

A good empirical choice for $x_{\max}$ has been found to be the value at which $\exp(-x_{\max}^\nu)=0.03$,

$$x_{\max}=[-\log_e(0.03)]^{1/\nu}\approx(3.5)^{1/\nu}. \tag{9}$$

The interpolation points still need to be distributed across the approximation interval. A reliable guide here is the observation that portions of a function with a large derivative generally require closer sampling to produce an accurate representation of the function's behavior. Since $\exp(-x^\nu)$ has a considerably larger derivative near $x=0$ than near $x=x_{\max}$, an accurate approximation will call for interpolation points that are more closely spaced near the origin. Furthermore, they should crowd closer together as ν decreases. A satisfactory way to achieve this is with interpolation points of the form

$$x_i=\left(\frac{i}{N}\right)^\beta x_{\max} \quad \text{for } i=1,\ldots,N, \tag{10a}$$

with an exponent

$$\beta=6-4.1(\nu-0.2). \tag{10b}$$

D. Pole splitting

Although $P(x)$ and $\exp(-x^\nu)$ agree at the interpolation points, they could still differ significantly elsewhere. That will certainly be the case if any pole falls within the approximation interval. Fortunately, when this happens it is possible to recover a good approximation through *pole splitting*. This is a process of replacing the pole $p>0$ with a pair of rotated poles $(pe^{i\delta},pe^{-i\delta})$ and simultaneously introducing a new zero to keep $M-L$ unchanged. In the pole-zero form of $P(x)$, this procedure is equivalent to the substitution

$$\frac{1}{1-x/p}\rightarrow\frac{1-x/z}{(1-x/pe^{i\delta})(1-x/pe^{-i\delta})}, \tag{11a}$$

with the added zero given by

$$z=\frac{p}{2\cos\delta-1}, \tag{11b}$$

an expression that emerges when the Taylor expansions for the two sides of Eq. (11a) are required to match through first order. The pole rotation angle δ is essentially arbitrary

though $\delta > 60°$ is advisable to prevent $z > 0$ (see the following section). Experience supports $\delta = 90°$ as a good choice.

E. Zero splitting

Since $\exp(-x^\nu)$ is positive definite, a suitable $P(x)$ ought to have no zeros in the approximation interval. Should one exist, it can be handled through a procedure similar to pole splitting: the zero is replaced with a pair of rotated zeros and an additional pole is introduced to leave $M-L$ unaltered. This amounts to the substitution

$$1 - x/z \rightarrow \frac{(1 - x/ze^{i\phi})(1 - x/ze^{-i\phi})}{1 - x/p}, \tag{12a}$$

where

$$p = \frac{z}{2\cos\phi - 1} \tag{12b}$$

in the pole-zero form of $P(x)$. Again, $\phi > 60°$ is necessary to prevent $p > 0$, and experience favors $\phi = 90°$.

F. Polynomial orders

Since $\exp(-x^\nu)$ decays with increasing x, $P(x)$ must do the same. Hence the restriction $M - L > 0$. In practice, it has been found that $M - L = 3$ yields good results. As would be expected, the approximation becomes more accurate for larger values of L and M—up to a point. It has been found that the approximation is apt to worsen beyond $M = 12$, due to numerical ill-conditioning in Eq. (7).

III. EXAMPLES

The examples below explore $2.4 \leqslant \gamma \leqslant 3.9$, the range of greatest practical interest for ocean acoustics. They examine the accuracy of the rational-function approximation by comparing it, where possible, to existing solutions (the two-series approximation[1,7] for $\gamma > 3$, the exact solution for $\gamma = 3$). The purpose is to determine the range of α throughout which the approximation matches the correct value to within a specified error. The error is expressed in decibels, and thus also represents the resulting error in any scattering strength calculated using the approximate $I(\alpha)$.

Each example demonstrates that, for the stated γ, the rational-function technique can yield a valid approximation of $I(\alpha)$ for all α above some small minimum value, $\alpha_{\min}$. To help assess the practical significance of $\alpha_{\min}$ for sea-floor scattering, we use a monostatic (e.g., backscatter) geometry to relate α to the grazing angle θ. The resulting $\theta_{\min}$ is the minimum grazing angle at which the error is within acceptable limits. As noted above, the α-to-θ mapping also depends on frequency. The examples are all done at 3.5 kHz, but we find the outcomes to be representative of results obtained for frequencies from 1 to 20 kHz.

There are four examples. The first two are in the range $3 < \gamma < 4$ where the two-series method yields accurate approximations of $I(\alpha)$ that can be used to benchmark the rational-function method. For the third, $\gamma = 3$ so that the exact solution can be used for validation. The final example has $2 < \gamma < 3$.

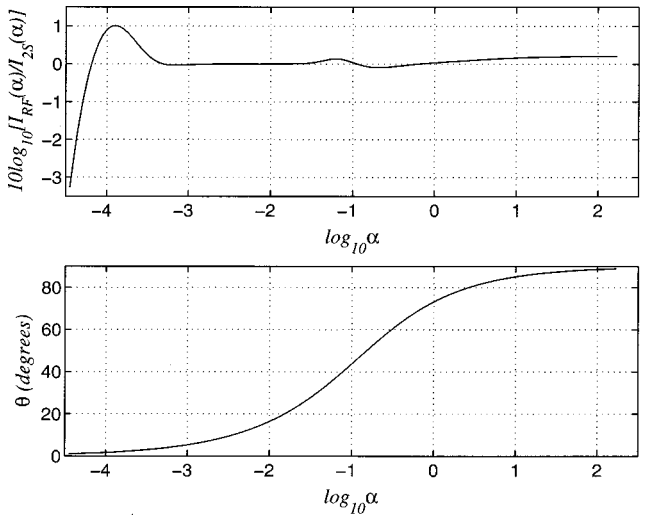


FIG. 1. Illustration of the quality of the approximation in the first example ($\gamma = 3.8$ with $L,M = 9,12$). (Top) Decibel comparison of the scattering integral $I(\alpha)$ by the rational-function method and the same integral by the two-series method. (Bottom) Grazing angle θ versus the parameter α.

A. Spectral exponent $\gamma = 3.8$

We first consider a spectral exponent in the range $\gamma > 3.5$, where it is particularly important to have accurate results because small-slope theory, which relies on $I(\alpha)$, yields scattering strengths that differ substantially from perturbation predictions.[4,5,7,11] This example uses $\gamma = 3.8$, a relatively large spectral exponent value that has been extracted from ocean surface scattering strength data.[11] The quality of the approximation for $L,M = 9,12$ is shown in Fig. 1. The top half of that figure shows the decibel difference between the rational-function approximation, $I_{\mathrm{RF}}(\alpha)$, and the two-series approximation, $I_{2\mathrm{S}}(\alpha)$. A logarithmic scale is used for the α axis to allow a more detailed examination of the $\alpha \ll 1$ region. The bottom half of the figure shows the relation between the grazing angle θ and the parameter α throughout the entire physical range, $0° \leqslant \theta \leqslant 90°$. The figure as a whole shows that the rational-function approximation agrees with the two-series approximation to within 1 dB for $\theta_{\min} \approx 2°$ and to within 3 dB for grazing angles down to $\theta_{\min} \approx 1°$. The result is a typical one. The rational-function method works very well here.

B. Spectral exponent $\gamma = 3.2$

Consider now $\gamma = 3.2$, a figure typical of values used to model bottom scatter from sediments[12] and from limestone bottoms.[5,7,13] The quality of the approximation for $L,M = 9,12$ is indicated by Fig. 2, which is identical in format to Fig. 1 except that the scope of the horizontal axis is restricted to $\alpha < 0.1$ to focus on small θ. For this case, there is agreement to within 1 dB for $\theta_{\min} \approx 8°$ and to within 3 dB at grazing angles down to $\theta_{\min} \approx 3°$.

To see how pole and zero splitting operate to recover a good approximation to $\exp(-x^\nu)$, consider the low-order case $L,M = 3,6$. With these polynomial orders, Eqs. (9), (10) yield $\beta = 4.36$, $x_{\max} = 8.093\,47$, and the interpolation points $x_1, \ldots, x_9$ shown in Fig. 3. Solving Eq. (7) then produces an approximation with these poles and zeros:

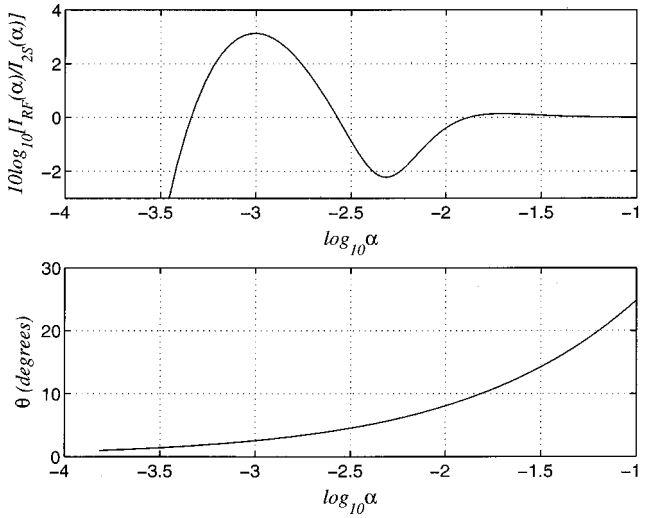


FIG. 2. Illustration of the quality of the approximation in the second example ($\gamma=3.2$ with $L,M=9,12$). (Top) Decibel comparison of the scattering integral $I(\alpha)$ by the rational-function method and the same integral by the two-series method. (Bottom) Grazing angle θ versus the parameter α.

$$p_1,\ldots,p_6=-0.837\,98\pm 6.633\,86i,\mathbf{1.855\,37},$$
$$-0.964\,65,-0.078\,77,-0.001\,59,$$
$$z_1,z_2,z_3=\mathbf{1.856\,76},-0.097\,57,-0.001\,65\,.$$

Each group has a member (**bold**) in the approximation interval, the combined effect of which is the singularity at $x\approx 1.85$ in Fig. 3(a). The resulting approximation is good except near this singularity. Splitting the pole removes the singularity but yields an inaccurate result because $P(x)$ still has a zero in the approximation interval, Fig. 3(c). Subsequent splitting of that zero yields the final approximation, Fig. 3(e), with the following poles and zeros:

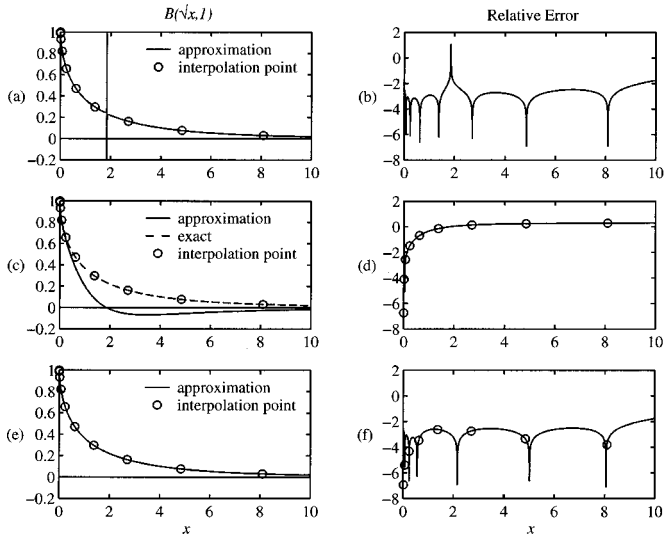


FIG. 3. Illustration of the effects of splitting poles and zeros in the second example ($\gamma=3.2$ with $L,M=3,6$). (Top) Left: the rational-function approximation generated by the direct solution of Eq. (7) without pole or zero splitting. Right: the relative error $R(x)$ for that approximation. Note the singularity in (a) and (b) due to the pole and zero within the approximation interval. The relative errors at the interpolation points are below -8 and occur at the sharp cusps. (Middle) The same quantities after pole splitting. (Bottom) The same quantities after both pole and zero splitting.

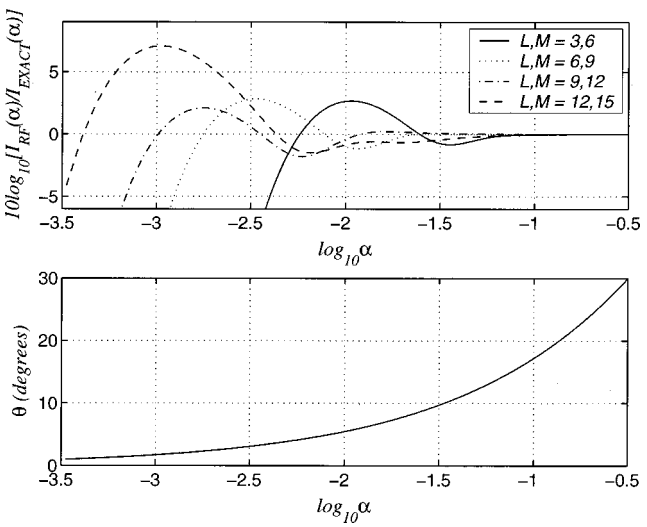


FIG. 4. Illustration of the quality of the approximation in the third example ($\gamma=3.0$). (Top) Decibel comparison of the scattering integral $I(\alpha)$ by the rational-function method and the exact solution. Results for four polynomial orders L,M are plotted. (Bottom) Grazing angle θ versus the parameter α.

$$p_1,\ldots,p_8=-0.837\,98\pm 6.633\,86i,\pm 1.855\,37i,$$
$$-0.964\,65,-0.078\,77,-0.001\,59,$$
$$-1.856\,76,$$
$$z_1,\cdots,z_5=\pm 1.856\,76i,-0.097\,56,-0.001\,65,$$
$$-1.855\,37\,.$$

The quality of the approximation at each stage in this process can be gauged by the *relative error*

$$R(x)\doteq\log_{10}\left|\frac{P(x)-\exp(-x^{\nu})}{\exp(-x^{\nu})}\right|, \tag{13}$$

whose negative, $-R(x)$, may be interpreted as the number of significant digits of agreement for the approximation $P(x)\approx B(\sqrt{x},1)$. The relative errors for the approximations before and after pole and zero splitting are shown in Figs. 3(b), (d), (f). In particular, Fig. 3(f) illustrates the relative error at the final stage in this example: $R(x)<-2$ for $0<x<9$, indicating agreement to two decimal places throughout the approximation interval.

C. Spectral exponent $\gamma=3.0$

For the third example, consider $\nu=1/2$, the only case with an exact, closed-form solution[14]

$$I(\alpha)=\frac{\alpha}{(1+\alpha^2)^{3/2}}. \tag{14}$$

TABLE I. Minimum grazing angles for the third example ($\gamma=3.0$).

Numerator order, L	Denominator order, M	$\theta_{\min}$ for 3 dB	$\theta_{\min}$ for 1 dB
3	6	4°	8°
6	9	3°	6°
9	12	2°	5°
12	15	4°	7°

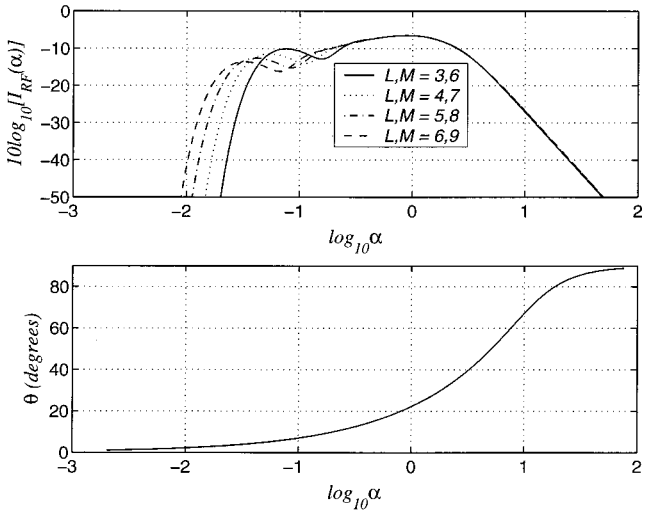


FIG. 5. (Top) Scattering integral $I(\alpha)$, in decibels, for the fourth example ($\gamma=2.6$) by the rational-function method. Results for four polynomial orders L,M are plotted. (Bottom) Grazing angle θ versus the parameter α.

The quality of the rational-function approximation can be judged from the decibel comparison with the exact solution, which is plotted in Fig. 4. Results from four polynomial orders—$L,M=3,6$; 6,9; 9,12; and 12,15—are included. Table I summarizes the minimum grazing angles for agreement within 3 dB and 1 dB. It is clear from this that there is very good agreement down to low grazing angles. Furthermore, increasing the polynomial orders of the rational function improves the accuracy of the approximation (i.e., reduces $\theta_{\min}$), but only up to a certain point, $M=12$ in this case. Beyond that point, numerical effects [namely ill-conditioning in Eq. (7)] actually make the approximation worse.

D. Spectral exponent $\gamma=2.6$

As a final example, consider $\nu=0.3$, a value toward the low end of the practical range. This corresponds to the spectral exponent $\gamma=2.6$, a value that has been extracted from high-resolution bathymetry measurements on scarps of the Mid-Atlantic Ridge.[5] Again, a series of four polynomial orders L,M is considered. The values of $I(\alpha)$ resulting from the rational-function approximations are plotted in Fig. 5 (alone, in this case, because the exact solution and two-series approximation are both unavailable for $\gamma<3$). Figure 6 is an enlargement of a portion of this same plot chosen to display the development of error in the rational-function approximation at low grazing angles. Table II contains the lower

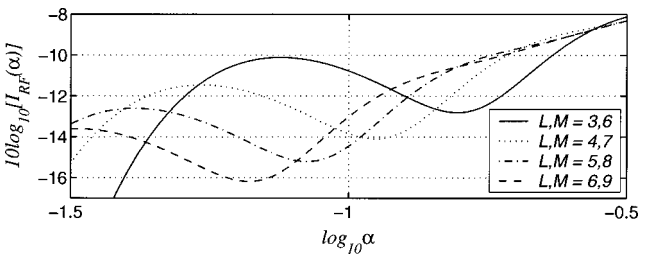


FIG. 6. Detail from Fig. 5 showing the small-θ details.

TABLE II. Minimum grazing angles for the fourth example ($\gamma=2.6$).

Numerator order, L	Denominator order, M	$\theta_{\min}$ for 1 dB
3	6	11°
4	7	9°
5	8	8°
6	9	7°

bounds on θ for a 1 dB error obtained by comparing each curve in the upper half of the figure with the monotonic θ-vs-α curve in the lower half. Throughout this range of increasing polynomial orders, there is a steady reduction of $\theta_{\min}$. For larger L and M, this trend will ultimately break down, as in the previous example, though we did not pursue that for this example.

IV. SUMMARY AND CONCLUSIONS

This paper documents a new algorithm for the numerical calculation of Eq. (1a), an integral that occurs widely in the bistatic theory of scattering from random rough interfaces. The algorithm is based on a rational-function approximation to the exponential factor in the integrand for the case $\alpha=1$, and leads directly to a closed-form expression for $I(1)$. A scaling property provides a means of extending this result to arbitrary α. In this form, the procedure is subject to instabilities due to any poles or zeros of the rational-function expression that happen to lie on the integration path. Original techniques are introduced for recovering a suitably accurate approximation to $I(\alpha)$ by splitting these poles and zeros into complex-conjugate pairs that are rotated away from the real axis. Examples of the resulting algorithm's operation in ocean-acoustic scattering scenarios are presented, complete with error analysis.

These examples (and others with different roughness spectral exponents and frequencies, which are not included in this paper) serve to validate the rational-function-based approximation of the scattering integral down to grazing angles of only a few degrees. Although the approximation does begin to fail at very small θ, there are good reasons why this should not be considered a problem. First, bottom scattering strengths can only be measured down to grazing angles of a few degrees with contemporary experimental techniques. Second, the scattering theories in which $I(\alpha)$ arises are local in character. They do not include diffraction or shadowing and thus are not valid a very low grazing angles. Third, predicted surface scattering strengths always vanish quite rapidly with grazing angle, routinely dropping to 50 dB or so below their peak values by the time θ reaches four or five degrees.[4] Hence, the consequences of the loss of some accuracy at these low angles are relatively minor. For these reasons, the computational algorithm described here is usable at most grazing angles of experimental and theoretical interest. For applications requiring greater accuracy at small grazing angles, Ref. 5 merges the algorithm developed here with the small-α series of Ref. 7 to produce an efficient numerical algorithm for computing $I(\alpha)$ for all θ and arbitrary γ.

ACKNOWLEDGMENTS

The authors are grateful for numerous useful discussions with Dr. Daniel Wurmser and Dr. Roger C. Gauss of the Naval Research Laboratory. This work was supported by the Office of Naval Research.

[1] R. Dashen, F. S. Henyey, and D. Wurmser, "Calculations of acoustic scattering from the ocean surface," J. Acoust. Soc. Am. **88**, 310–323 (1990).

[2] D. R. Jackson, D. P. Weinbrenner, and A. Ishimaru, "Application of the composite roughness model to high-frequency bottom backscattering," J. Acoust. Soc. Am. **79**, 1410–1422 (1986).

[3] R. Dashen and D. Wurmser, "Approximate representations of the scattering amplitude," J. Math. Phys. **32**, 986–996 (1991).

[4] D. Wurmser, R. F. Gragg, and R. C. Gauss, "Calculations of acoustic scattering from an elastic ocean bottom," J. Acoust. Soc. Am. **104**, 1809 (1998).

[5] R. F. Gragg, D. Wurmser, and R. C. Gauss, "Small-slope scattering from rough elastic ocean floors: General theory and computational algorithm," J. Acoust. Soc. Am. (to be published).

[6] Since the spectrum cannot actually diverge as $k \to 0$, the power-law tail $S(k) \propto k^{-\gamma}$ really should be augmented with a cutoff of some type at some small wave number k_c. However, the physical mechanism at work is Bragg scattering at wave number k_B, and it is tacitly assumed that k_B is well above k_c, so that the cutoff's existence can be neglected.

[7] R. F. Gragg and D. Wurmser, "Scattering from rough elastic ocean floors," *Proceedings of the Fifth European Conference on Underwater Acoustics*, Volume 2, edited by M. E. Zakharia *et al.* (European Commission, Luxembourg, 2000), pp. 1177–1182.

[8] A more general version would be needed if repeated poles occurred, but this has never happened in any numerical implementation of the algorithm reported here. It may be viewed as a pathological case that is not encountered in practice.

[9] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, New York, 1995), Chap. 13, p. 434.

[10] B. P. Lathi, *Signals, Systems and Communication* (Wiley, New York, 1965).

[11] R. C. Gauss and J. M. Fialkowski, "A broadband model for predicting bistatic surface scattering strengths," *Proceedings of the Fifth European Conference on Underwater Acoustics*, Volume 2, edited by M. E. Zakharia *et al.* (European Commission, Luxembourg, 2000), pp. 1165–1170.

[12] K-Y. Moravan and D. R. Jackson, "Model/Data comparison for bottom backscatter in the North Atlantic at low frequencies," APL-UW TR 9304, Applied Physics Laboratory, University of Washington, February, 1993.

[13] D. M. Drumheller and R. F. Gragg, "Numerical evaluation of an integral found in the theory of scattering from rough interfaces," Naval Research Laboratory Memorandum Report No. 7140-00-8436, March 17, 2000.

[14] Although Eq. (1) is also integrable in closed form for $\nu = 1$, Eq. (2) implies that $\alpha = \infty$ in that case.

An effective density fluid model for acoustic propagation in sediments derived from Biot theory

Kevin L. Williams[a)]
Applied Physics Laboratory, College of Ocean and Fishery Sciences, University of Washington, 1013 N.E. 40th Street, Seattle, Washington 98105

(Received 6 April 2001; revised 6 August 2001; accepted 27 August 2001)

In this paper we present an acoustic propagation model that approximates a porous medium as a fluid with a bulk modulus and effective density derived from Biot theory. Within the framework of Biot theory it is assumed here that the porous medium has low values of frame bulk and shear moduli relative to the other moduli of the medium and these low values are approximated as zero. This leads to an effective density fluid model. It is shown that, for saturated sand sediments, the dispersion, transmission, reflection, and in-water backscattering predicted with this effective density fluid model are in close agreement with the predictions of Biot theory. In this agreement we demonstrate that the frame bulk and shear moduli play only a minor role in determining several aspects of sand acoustics. Thus, for many applications the effective density fluid model is an accurate alternative to full Biot theory and is much simpler to implement. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1412449]

PACS numbers: 43.20.Fn, 43.20.Hq, 43.30.Hw, 43.30.Ma [DLB]

I. INTRODUCTION

Studies of acoustic propagation[1–7] within, penetration[8–15] into, and scattering[16–21] from sand sediments have a long history (Refs. 1–21 represent starting points into the literature). Constitutive equations used in studies of the interaction of acoustic energy with sand include fluid, viscoelastic, and poroelastic. The poroelastic model originally developed by Biot[22,23] has received more attention lately due to the possibility that penetration at shallow grazing angles could be explained by the existence of the Biot slow wave.[10] Though recent research indicates that scattering from interface roughness is the more dominant mechanism for such shallow grazing angle penetration,[12–14,24,25] Biot theory has been successful in explaining dispersion[15] and has been shown to predict different levels of scattering[21] from the rough sand interface than found in previous fluid[18] and viscoelastic[19,20] models. These results in part motivated the research summarized here.

In this paper we present an acoustic propagation model that approximates a fluid-saturated porous medium as a fluid with a bulk modulus and effective density derived from Biot theory. It is shown that, for sand sediments, the dispersion, transmission, reflection, and scattering predicted with this effective density fluid model (EDFM) are very close to the predictions of Biot theory. The EDFM thus gives insight into the environmental parameters that are most important in determining the acoustic behavior of sand. It can also be an alternative, that is much simpler to implement, to using the full Biot model.

In Sec. II we summarize the derivation of the EDFM. In Sec. III dispersion results for the EDFM and Biot models are compared. Then the reflection, transmission, and backscattering derived using a previous fluid model (FM),[18] the EDFM, and the Biot model are examined for this same sand. Section IV is a brief summary.

II. DETERMINATION OF AN EFFECTIVE DENSITY FLUID MODEL FROM BIOT EQUATIONS

In Sec. II A we summarize the equations of motion for the Biot model in a notation based primarily on Ref. 16, but with an $\exp(-i\omega t)$ assumed time dependence for the wave solutions. In Sec. II B we then present the EDFM that results from setting the frame bulk modulus K_b and shear modulus μ to zero (in the context of a sediment these are the moduli associated with the assemblage of the sediment particles[16]). The behavior of the effective density as a function of frequency is further examined in Sec. II C.

A. Biot model

Stoll[16] uses potentials $(\Phi_s, \Phi_f, \Psi_s, \Psi_f)$ defined in terms of the displacement vectors of the skeletal frame ($\mathbf{u}$) and the water ($\mathbf{U}$) as follows:

$$\mathbf{u} = \nabla\Phi_s + \nabla\times\Psi_s,$$
$$\beta(\mathbf{u}-\mathbf{U}) = \nabla\Phi_f + \nabla\times\Psi_f, \tag{1}$$

where β is the sediment porosity. Biot's equations for the scalar potentials are then given as (using a plane wave solution $\exp[i(\mathbf{k}\cdot\mathbf{x}-\omega t)]$)

$$(-k^2H)\Phi_s + k^2C\Phi_f = (-\omega^2)\rho\Phi_s + \rho_f\omega^2\Phi_f, \tag{2}$$

$$(-k^2C)\Phi_s + k^2M\Phi_f = (-\omega^2)\rho_f\Phi_s + \frac{\omega^2\alpha\rho_f}{\beta}\Phi_f + \frac{i\omega F\eta}{\kappa}\Phi_f. \tag{3}$$

In Eqs. (2) and (3), k is the acoustic wave number, α is the structure constant or tortuosity,[26] η is the pore fluid viscosity,

a)Electronic mail: williams@apl.washington.edu

κ is the permeability, ρ_f is the pore fluid mass density, ρ_s is the sediment particle mass density, and

$$
\begin{aligned}
&k=\frac{\omega}{c}, \quad \omega=2\pi f,\\
&\rho=\beta\rho_f+(1-\beta)\rho_s,\\
&H=[(K_r-K_b)^2/(D-K_b)]+K_b+(4\mu)/3,\\
&C=[K_r(K_r-K_b)]/(D-K_b),\\
&M=K_r^2/(D-K_b),\\
&D=\dot{K}_r[1+\beta(K_r/K_f-1)],
\end{aligned}
\tag{4}
$$

where f is the frequency, c is the speed of sound, ρ is the total mass density, K_r is the bulk modulus of individual sediment grains, and K_f is the bulk modulus of the pore fluid. The parameter F represents the deviation from Poiseuille flow as frequency increases.[23] Assuming cylindrical pores, Biot derived an expression for F that in the present notation is

$$
F(\epsilon)=\frac{\frac{\epsilon}{4}T(\epsilon)}{1-\frac{2i}{\epsilon}T(\epsilon)}, \tag{5}
$$

with

$$
T(\epsilon)=\frac{(-\sqrt{i})J_1(\epsilon\sqrt{i})}{J_0(\epsilon\sqrt{i})}, \tag{6}
$$

where J_1 and J_0 are cylindrical Bessel functions and

$$
\epsilon=a\sqrt{\frac{\omega\rho_f}{\eta}}. \tag{7}
$$

The parameter a is the "pore size parameter." Johnson *et al.*[27] examined the problem of a pressure gradient applied across a porous sample and found that for a porous media consisting of nonintersecting canted tubes the relationship for a (that is, the pore radius for the canted tube model),

$$
a=\sqrt{\frac{8\alpha\kappa}{\beta}}, \tag{8}
$$

holds. Furthermore, they examined alternate pore models as well as experimental results to conclude that "the relationship may be violated by a factor of 2 but not, apparently, a factor of 10." In the results examined here, Eq. (8) was used to determine a.

Using Eqs. (2) and (3), a fourth-order polynomial for the (complex) wave number k can be found. The solution gives two different wave numbers corresponding to the Biot fast and slow waves.[16] Both these waves show significant dispersion. The Biot fast wave dispersion will be shown in Sec. III A.

B. Limit when K_b, μ equal zero

Taking the $K_b=\mu=0$ limit of Biot theory is motivated by the fact that for sand sediments the frame and shear moduli are much lower than the other moduli.[16] Using Eq. (4), it is easy to show that, when K_b and μ are set to zero,

$$
H=C=M=\left(\frac{(1-\beta)}{K_r}+\frac{\beta}{K_f}\right)^{-1}. \tag{9}
$$

This result states that the compressibility (modulus^{-1}) is a linear function of the concentration of the particles in a suspension as assumed by Urick in Ref. 28. Using Eq. (9) in Eqs. (2) and (3) gives

$$
(-k^2H)\Phi_s+k^2H\Phi_f=(-\omega^2)\rho\Phi_s+\rho_f\omega^2\Phi_f, \tag{10}
$$

$$
(-k^2H)\Phi_s+k^2H\Phi_f=(-\omega^2)\rho_f\Phi_s+\frac{\omega^2\alpha\rho_f}{\beta}\Phi_f+\frac{i\omega F\eta}{\kappa}\Phi_f. \tag{11}
$$

Using Eqs. (10) and (11), a second-order equation for the wave number k can be found [as opposed to the fourth-order equation for Eqs. (2) and (3)]. This solution and the relation from Eq. (4) that $k=\omega/c$ can be used to find that

$$
k^2=\frac{\omega^2\rho_{\mathrm{eff}}(\omega)}{H} \tag{12}
$$

or

$$
c=\sqrt{\frac{H}{\rho_{\mathrm{eff}}(\omega)}}, \tag{13}
$$

where

$$
\rho_{\mathrm{eff}}(\omega)=\rho_f\left(\frac{\alpha(1-\beta)\rho_s+\beta(\alpha-1)\rho_f+\frac{i\beta\rho F\eta}{\rho_f\omega\kappa}}{\beta(1-\beta)\rho_s+(\alpha-2\beta+\beta^2)\rho_f+\frac{i\beta F\eta}{\omega\kappa}}\right), \tag{14}
$$

is a complex effective density.

Using Eqs. (11) and (12), Φ_f can be found in terms of Φ_s,

$$
\Phi_f=\left(\frac{\rho_{\mathrm{eff}}(\omega)-\rho_f}{\rho_{\mathrm{eff}}(\omega)-\frac{\alpha\rho_f}{\beta}-\frac{iF\eta}{\omega\kappa}}\right)\Phi_s. \tag{15}
$$

Then, using Eqs. (14), (15), and (10), the equation,

$$
k^2H\Phi=\omega^2\rho_{\mathrm{eff}}(\omega)\Phi, \tag{16}
$$

can be derived where $\Phi=\Phi_s+\Phi_f$. Equation (16) has the form of a fluid equation of motion (for a plane wave) but with a density that is a function of frequency [Eq. (14)] and a modulus H that is determined by the moduli of the suspension components [Eq. (9)]. Equations (9), (14), and (16) are the defining equations of the EDFM.

C. Examination of effective density

It is useful to examine the effective density of Eq. (14) in some detail and relate it to previous results. In the low-frequency limit, Eq. (14) becomes

$$
\rho_{\mathrm{eff}}(\omega\to 0)=\rho. \tag{17}
$$

That is, the effective density becomes equal to the total mass density, as given in Eq. (4). This low-frequency limit is the

TABLE I. Parameters assumed in calculating Biot results. The results for the effective density fluid model (EDFM) were calculated using the same parameters, except that K_b and μ were set to zero. Note that K_b and μ are complex.

Parameter	Value
Porosity (β)	0.40
Mass density of sand grains (ρ_s)	2650 kg/m^3
Mass density of water (ρ_f)	1000 kg/m^3
Bulk modulus of sand grains (K_r)	3.6×10^{10} Pa
Bulk modulus of frame (K_b)	$4.4\times10^{7}-2.0\times10^{6}i$ Pa
Shear modulus of frame (μ)	$2.6\times10^{7}-1.25\times10^{6}i$ Pa
Bulk modulus of water K_f	2.25×10^{9} Pa
Viscosity (η)	0.001 kg/m·s
Permeability (κ)	1.0×10^{-10} m^2
Tortuosity (α)	1.25

density used in the FM scattering model. Thus, it may be anticipated that in the reflection, transmission, and scattering results of the next section the EDFM and FM should agree at low frequencies.

In the high-frequency limit the density becomes

$$\rho_{\text{eff}}(\omega\to\infty)=\rho_f\left(\frac{\alpha(1-\beta)\rho_s+\beta(\alpha-1)\rho_f}{\beta(1-\beta)\rho_s+(\alpha-2\beta+\beta^2)\rho_f}\right). \quad (18)$$

This expression is also obtained from multiple scattering theory in the case of ordered cubic suspensions,[6] assuming a nonviscous fluid. [In Ref. 6 it is shown that Eq. (18) is not generally valid for disordered media. However, for the density contrast between sand and water, Eq. (18) is an accurate approximation to the effective density determined via multiple scattering theory.] This same result [Eq. (18)] is implicit in the high-frequency limit to the fast wave speed given in Eq. (7b) of Ref. 5.

The parameters given in Table I were used to further examine the behavior of the effective density as a function of frequency (the parameters used here are in the range of those quoted in Table I of Ref. 10). Figure 1 is a plot of the real and imaginary parts of the effective density as a function of frequency from 10 Hz to 100 kHz. Also shown are the low- and high-frequency limits of Eqs. (17) and (18). The physical reason for the transition between these two values can be understood from Biot theory.[23] At low frequencies the pore size parameter is much smaller than the viscous skin depth[29] and there is no relative motion between the grains and the water. At high frequencies the skin depth is much smaller than the pore size parameter (typical pore radius) and it is the tortuosity α that determines the relative motion of the fluid and solid [as seen by the high-frequency limit of Eq. (15)]. In this regard, note that in the limit that tortuosity goes to infinity Eq. (17) becomes true for all frequencies, that is, the fluid and solid are always moving together. The FM's use of the total mass density is thus equivalent to assuming that α equals infinity.

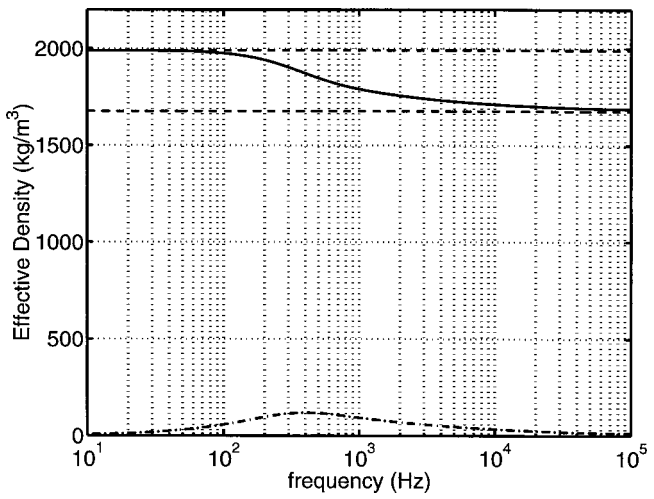


FIG. 1. The real (solid curve) and imaginary (dot–dashed curve) parts of the effective density given in Eq. (14) for the parameters in Table I. Also plotted are the low (top dashed line) and high- (bottom dashed line) frequency limits of Eq. (14) as given in Eqs. (17) and (18), respectively.

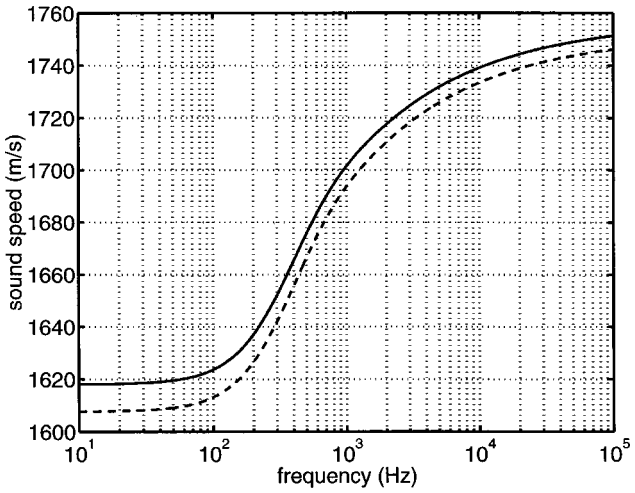


FIG. 2. The sound speed as a function of frequency calculated using the Biot model (solid curve) and the effective density fluid model (dashed curve) for the parameters in Table I.

III. COMPARISON OF RESULTS FOR A SAND SEDIMENT—PREVIOUS FLUID MODEL, EFFECTIVE DENSITY FLUID MODEL, AND FULL BIOT MODEL

The question to be examined here is whether the EDFM is a good approximation to a Biot model for at least some aspects of the acoustic behavior of a sand sediment. Setting K_b and $\boldsymbol{\mu}$ to zero to obtain the EDFM implies that the slow wave and shear wave do not exist in this model. However, many times it is the behavior of the fast wave within the sediment or the behavior of the acoustic wave in the water that is of interest. Within this section, dispersion results are presented for the EDFM and the Biot fast wave, and reflection, transmission, and backscattering results are presented using the FM, the EDFM, and the Biot model. Again, the parameter set used is that given in Table I.

Figures 2 and 3 show the sound speed and attenuation as a function of frequency for both the Biot model [solid lines—Eqs. (2) and (3) were used to derive the dispersion equation] and the EDFM [dashed lines—Eq. (13)]. Both models produce similar dispersion in sound speed, however, the EDFM predicts speeds that are about 11 m/s slower at the lowest frequencies shown and about 5 m/s slower at the highest frequencies. The average sound speed error is about

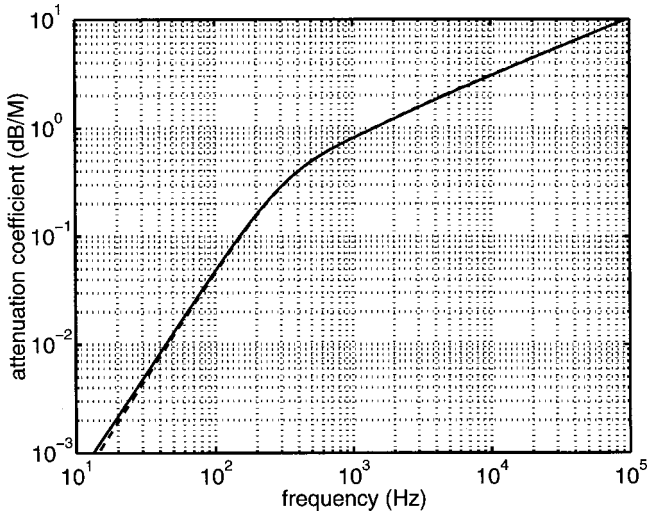


FIG. 3. Attenuation as a function of frequency calculated using the Biot model (solid curve) and the effective density fluid model (dashed curve) for the parameters in Table I.

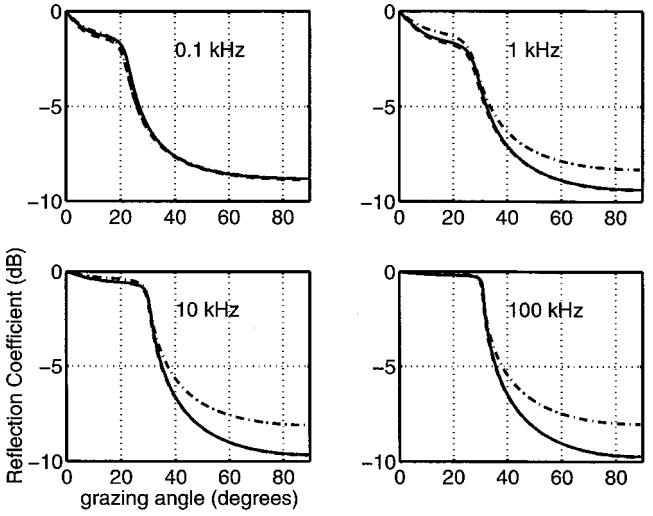


FIG. 4. The reflection coefficient as a function of grazing angle at four different frequencies for the FM (dot–dashed curves), the EDFM (dashed curves), and the Biot model (solid curves).

0.5% over the frequency range plotted. The bulk and shear moduli of the frame obviously increase the sound speed slightly. The effect of the larger difference in sound speed at lower frequencies will be seen in the reflection, transmission, and scattering results that follow. (It is possible to adjust the grain modulus K_r in the EDFM upward slightly and get very close agreement to the Biot dispersion curve, but this is not done here.)

The attenuation predicted by the EDFM is slightly lower at the lowest frequencies plotted. This discrepancy increases as the frequency is reduced below the range plotted due to internal frictional effects of the frame (as indicated by the complex frame bulk and shear moduli in Table I). The attenuation due to frame moduli varies linearly with frequency, whereas the attenuation due to water viscosity in both the Biot and EDFM model varies as frequency squared up to around 1 kHz in Fig. 3. Thus, at low enough frequencies the frame attenuation will begin to become important.

Figure 4 shows the reflection coefficient as a function of angle for four different frequencies (0.1, 1, 10, 100 kHz) for the FM[18] (dot–dashed curves), the EDFM (dashed curves), and the Biot[21] model (solid curves). In addition to the total mass density ρ, the FM model needs measured values of sound speed and attenuation as input, and in this and all subsequent plots the values used were those determined by the dispersion curve and attenuation of the EDFM model. Also, all EDFM calculations were carried out by using the FM formulas, but replacing the FM density by the effective density of Eq. (14). Thus, the only difference between the FM and EDFM calculations is the density they use. As anticipated in the previous section, all three models are in close agreement at 0.1 kHz. This is because the water is locked to the grains at low frequencies and thus the total mass density used by the FM is a close approximation to the effective density [Eq. (17)]. The biggest difference is near the critical angle (near 20°) and is due to the difference in sound speed between both the fluid models and the Biot model (Fig. 2).

At the higher frequencies the EDFM and Biot models actually come into closer agreement. This is due to the fact that the difference between the sound speeds of the two models becomes smaller as frequency is increased. At the same time the FM deviates further from the other two models because the total mass density is not a good approximation to the effective density at these frequencies (Fig. 1). Near normal incidence the difference between the FM and either Biot or EDFM is about 2 dB at 100 kHz.

This same general trend is seen in Fig. 5 for the transmission coefficient. That is, all three models are close at 0.1 kHz, however, Biot and EDFM become closer as the frequency is increased while the FM deviates from the other two models. The procedure used to calculate a transmission coefficient for the Biot model was to add the transmission coefficients found in Ref. 21 for the scalar potentials Φ_{1f} and Φ_{1s} associated with the propagation of the fast wave. Furthermore, the transmission coefficient for the fluid models derived from Ref. 18 is in terms of pressure not scalar potential, so it was divided by the density ratio between the

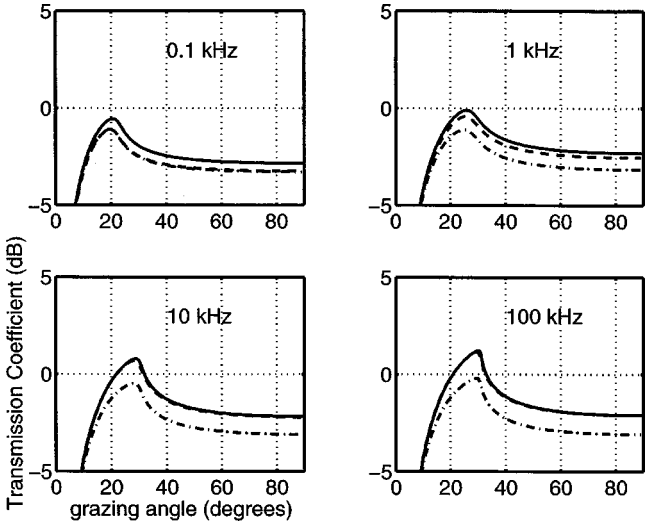


FIG. 5. The transmission coefficient as a function of grazing angle at four different frequencies for the FM (dot–dashed curves), the EDFM (dashed curves), and the Biot model (solid curves).

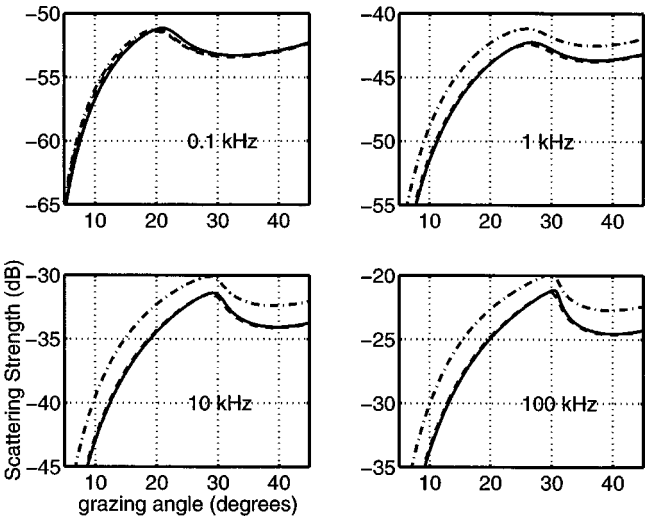


FIG. 6. The scattering strength as a function of grazing angle at four different frequencies for the FM (dot–dashed curves), the EDFM (dashed curves), and the Biot model (solid curves).

sediment (effective density for EDFM and total mass density for FM) and water before plotting in Fig. 5. The dB error near normal incidence (about 1 dB at 100 kHz) is less than for reflection because more energy is transmitted into the sediment than reflected. Thus, a 2 dB change in reflection gives less than a 2 dB change in transmission.

Figure 6 shows the backscattering strength as a function of grazing angle using Ref. 18 for the FM and EDFM and Ref. 21 for the Biot model. The surface roughness in these models[18,21] is given in terms of a roughness spectrum with two parameters; the spectral strength w_2, and the spectral exponent γ_2. For Fig. 6, $w_2 = 0.000\,05\,\mathrm{m}^4$ and $\gamma_2 = 3.05$. Again the EDFM is seen to be an accurate approximation to the Biot model result, whereas the FM differs by close to 3 dB in some cases (e.g., in the vicinity of 15° grazing angle at 100 kHz).

IV. SUMMARY

The accuracy of the EDFM relative to Biot theory confirms that (for typical environmental parameters used) the frequency dependence of sound speed and attenuation (over the range of 10 Hz to 100 kHz) seen in the Biot model for sand is almost entirely due to changes in the motion of the water relative to the sand grains. The agreement between Biot theory and EDFM demonstrates that the bulk frame and shear moduli play a minor role in reflection, transmission, and in-water backscattering. Thus, for many applications the EDFM is an accurate alternative to full Biot theory and is much simpler to implement.

The EDFM requires eight input parameters: the density and bulk modulus of sediment grains, density, bulk modulus, and viscosity of water, and the porosity, permeability, and tortuosity of the sediment. It is probably measurement uncertainties of grain bulk modulus, and sediment permeability and tortuosity that result in the greatest EDFM uncertainty.

The FM[18] assumes that the appropriate sediment density to be used in acoustic studies is the total mass density. In the EDFM and Biot model this is true only at low frequencies. (Note that as tortuosity goes to infinity the FM, EDFM, and Biot model are in close agreement at all frequencies. This limit is of potential interest for sediments where most pore channels are blocked.) Recent research on dispersion[15] and backscattering[24] have begun to point to the Biot model as a more accurate alternative to the FM. Much of the important physics responsible for this dispersion and backscattering within the Biot model is also present in the EDFM.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. The author acknowledges the many discussions with Darrell Jackson, Eric Thorsos, Dajun Tang, and Steve Kargl that helped clarify some of the important issues in sediment acoustics and led to this work.

[1] L. D. Hampton, "Acoustic properties of sediments," J. Acoust. Soc. Am. **42**, 882–890 (1967).
[2] E. L. Hamilton, "Prediction of in-situ acoustic and elastic properties of marine sediments," Geophysics **36**, 266–284 (1971).
[3] R. D. Stoll, "Acoustic waves in saturated sediments," in *Physics of Sound in Marine Sediments*, edited by L. Hampton (Plenum, New York, 1972), pp. 19–39.
[4] J. M. Hovem and G. D. Ingram, "Viscous attenuation of sound in saturated sand," J. Acoust. Soc. Am. **66**, 1807–1812 (1979).
[5] D. L. Johnson and T. Plona, "Acoustic slow waves and the consolidation transition," J. Acoust. Soc. Am. **72**, 556–565 (1982).
[6] L. M. Schwartz and D. L. Johnson, "Long-wavelength acoustic propagation in ordered and disordered suspensions," Phys. Rev. B **30**, 4302–4313 (1984).
[7] P. R. Ogushwitz, "Applicability of the Biot theory. II. Suspensions," J. Acoust. Soc. Am. **77**, 441–452 (1985).
[8] T. G. Muir, C. W. Horton, and L. A. Thompson, "Penetration of highly directional acoustic beams into sediments," J. Sound Vib. **64**, 539–551 (1979).
[9] F. B. Jenson and H. Schmidt, "Subcritical penetration of narrow gaussian beams into sediments," J. Acoust. Soc. Am. **82**, 574–579 (1987).
[10] N. P. Chotiros, "Biot model of sound propagation in water-saturated sand," J. Acoust. Soc. Am. **97**, 199–214 (1995).
[11] K. L. Williams, L. J. Satkowiak, and D. R. Bugler, "Linear and parametric array transmission across a water–sand interface—Theory, experiment, and observation of beam displacement," J. Acoust. Soc. Am. **86**, 311–325 (1989).
[12] E. I. Thorsos, D. R. Jackson, and K. L. Williams, "Modeling of subcritical penetration into sediments due to interface roughness," J. Acoust. Soc. Am. **107**, 263–277 (2000).
[13] A. Maguer, W. L. J. Fox, H. Schmidt, E. Pouliquen, and E. Bovio, "Mechanisms for subcritical penetration into a sandy bottom: Experimental and modeling results," J. Acoust. Soc. Am. **107**, 1215–1225 (2000).
[14] H. J. Simpson and B. H. Houston, "Synthetic array measurements of acoustic waves propagating into a water-saturates sandy bottom for a smoothed and roughened interface," J. Acoust. Soc. Am. **107**, 2329–2337 (2000).
[15] A. Maguer, E. Bovio, and W. L. J. Fox, "*In situ* estimation of sediment sound speed and critical angle," J. Acoust. Soc. Am. **108**, 987–996 (2000).
[16] R. D. Stoll and T. K. Kan, "Reflection of acoustic waves at a water–sediment interface," J. Acoust. Soc. Am. **70**, 149–156 (1981).
[17] D. R. Jackson, K. B. Briggs, K. L. Williams, and M. D. Richardson, "Tests of models for high-frequency seafloor backscatter," IEEE J. Ocean. Eng. **21**, 458–470 (1996).
[18] K. L. Williams and D. R. Jackson, "Bistatic bottom scattering: Model, experiments, and model/data comparison," J. Acoust. Soc. Am. **103**, 169–181 (1998).
[19] D. R. Jackson and A. N. Ivakin, "Scattering from elastic sea beds: First-order theory," J. Acoust. Soc. Am. **103**, 336–345 (1998).
[20] A. N. Ivakin and D. R. Jackson, "Effects of shear elasticity on sea bed scattering: Numerical examples," J. Acoust. Soc. Am. **103**, 346–354 (1998).
[21] K. L. Williams, J. M. Grochocinski, and D. R. Jackson, "Interface scat-

tering by poroelastic sea beds: First-order theory,'' J. Acoust. Soc. Am. (to be published).
[22] M. A. Biot, ''Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low frequency range,'' J. Acoust. Soc. Am. **28**, 168–178 (1956).
[23] M. A. Biot, ''Theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher frequency range,'' J. Acoust. Soc. Am. **28**, 179–191 (1956).
[24] E. I. Thorsos, K. L. Williams, D. R. Jackson, M. D. Richardson, K. B. Briggs, and D. Tang, ''An experiment in high-frequency sediment acoustics: SAX99,'' *Proceedings of the Institute of Acoustics Conference*, edited by T. G. Leighton, G. J. Heald, G. Griffiths, and H. D. Griffiths, South Hampton, United Kingdom, 2001.
[25] D. R. Jackson, K. L. Williams, E. I. Thorsos, and S. G. Kargl, ''High-frequency subcritical penetration into a sandy sediment,''submitted to IEEE J. Ocean Eng.
[26] B. Yavari and A. Bedford, ''Comparison of numerical calculations of two Biot coefficients with analytical solutions,'' J. Acoust. Soc. Am. **90**, 985–990 (1991).
[27] D. L. Johnson, J. Koplik, and R. Dashen, ''Theory of dynamic permeability and tortuosity in fluid-saturated porous media,'' J. Fluid Mech. **176**, 379–402 (1987).
[28] R. J. Urick, ''A sound velocity method for determining the compressibility of finely divided substances,'' J. Appl. Phys. **18**, 983–987 (1947).
[29] A. L. Fetter and J. D. Walecka, *Theoretical Mechanics of Particles and Continua* (McGraw-Hill, New York, 1980).

Simplified method to solve sound transmission through structures lined with elastic porous material

Joon-H. Lee[a] and J. Kim[b]
Structural Dynamics Research Laboratory, Department of Mechanical Engineering, University of Cincinnati, Cincinnati, Ohio 45221-0072

Han-J. Kim
Advanced Technology, ArvinMeritor Industries, Inc., Columbus, Indiana 47201

(Received 10 April 2001; revised 8 August 2001; accepted 16 August 2001)

An approximate analysis method is developed to calculate sound transmission through structures lined with porous material. Because the porous material has both the solid phase and fluid phase, three wave components exist in the material, which makes the related analysis very complicated. The main idea in developing the approximate method is very simple: modeling the porous material using only the strongest of the three waves, which in effect idealizes the material as an equivalent fluid. The analysis procedure has to be conducted in two steps. In the first step, sound transmission through a flat double panel with a porous liner of infinite extents, which has the same cross sectional construction as the actual structure, is solved based on the full theory and the strongest wave component is identified. In the second step sound transmission through the actual structure is solved modeling the porous material as an equivalent fluid while using the actual geometry of the structure. The development and validation of the method are discussed in detail. As an application example, the transmission loss through double walled cylindrical shells with a porous core is calculated utilizing the simplified method.
[DOI: 10.1121/1.1410967]

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

I. INTRODUCTION

While flat or curved thin elastic panels lined with porous material are found in many practical applications, an analysis of sound transmission through such structures remains as a very difficult task. Because the porous material has the solid phase (elastic frame) as well as the fluid phase (air contained in pores), three wave components are necessary to describe the acoustic response of the material.[1,2] Many theories with different degrees of approximation have been proposed to model the porous material, therefore to solve sound transmission through structures lined with porous material.[2–21] In some cases it is assumed that the solid constituent in the porous material is rigid, which allows only a single longitudinal wave in the material.[3–5] In some models the elasticity of the frame is considered but its shear deformation is neglected, thus allowing two longitudinal waves[6–8] in the material, which, however, does not allow oblique incident cases[6,9] to be handled. Bolton *et al.*[2] developed an analysis method, modeling the porous material based on the Biot's theory,[1] which considers the porous material as a homogeneous aggregate of the elastic frame and fluid contained in the pores. The theory[1,2] allows three waves in the porous material, two dilatational (also called longitudinal) waves, and one rotational (also called shear) wave, therefore can handle oblique waves. One of the dilatational waves is predominantly the elastic wave, therefore referred to as the frame wave, and the other is predominantly the acoustic wave, therefore referred to as the airborne wave. The rotational wave is predominantly the shear wave, or the secondary wave, as it is often referred to. This method will be referred to as the full method hereafter. The full method[2] was applied to calculate transmission losses (TLs) of elastic double panels of infinite extents lined with elastic porous liners. The TLs calculated for random incidences were compared with the measure TLs,[2] which showed quite good agreement.

Applying the full method to more general cases such as curved panels of finite extents becomes very difficult because of the complexity of the multiwave theory to describe the porous material. The finite element method (FEM) formulation of the multiwave theory was conducted by several researchers,[10–12,22,23] however its applications appear to be still limited to somewhat simple cases. This work was motivated by a desire to develop a simplified analysis method as an alternative to the full method so that the solutions to structures of complicated geometry lined with a layer can be obtained with relative ease and reasonable accuracy.

In this work, development of a simplified method for an analysis of the sound transmission through structures with porous material is explained at first. The validity of the method is checked by comparing the TLs obtained for double panel structures with the TLs obtained for the same structure but from the full method and measurements reported by Bolton *et al.*[2] Then, the method is applied to calculate TLs of a double-walled cylinder sandwiching a porous layer, which serves as an application example.

[a]Author to whom correspondence should be addressed. Electronic mail: jay.kim@uc.edu; telephone: (513) 556-6300.
[b]Currently with Motorola, Inc.

II. DEVELOPMENT OF SIMPLIFIED ANALYSIS METHOD

A. Review of the full theory

As a preliminary step to develop the simplified method to analyze sound transmission through structures with a porous liner, the method developed by Bolton *et al.*[2] based on Biot's theory,[1] which is referred to as the full method, is reviewed. It is noted that this full method is discussed only to the extent necessary to explain the simplified method obviously without claiming any contributions. Readers are suggested to read references[1,2,5,13,14] for more detailed discussions if necessary.

If the porous material is assumed a homogeneous and isotropic aggregate of the elastic frame and the fluid trapped in pores, its acoustic behavior can be described by the following two wave equations.[1]

$$\nabla^4 e_s + A_1 \nabla^2 e_s + A_2 e_s = 0, \tag{1}$$

$$\nabla^2 \varpi + k_t^2 \varpi = 0. \tag{2}$$

Equation (1) governs elastic dilatational waves, and Eq. (2) governs the rotational wave. In Eqs. (1) and (2), $e_s = \nabla \cdot \bar{u}$ is the solid volumetric strain, $\bar{u}$ being the displacement vector of the solid, $\varpi = \nabla \times \bar{u}$ is the rotational strain in the solid phase. Also, parameters A_1 and A_2 are

$$A_1 = \frac{\omega^2(\rho_{11}^* R - 2\rho_{12}^* Q + \rho_{22}^* P)}{(PR - Q^2)}, \tag{3}$$

$$A_2 = \frac{\omega^4[\rho_{11}^* \rho_{22}^* - (\rho_{12}^*)^2]}{(PR - Q^2)}, \tag{4}$$

where ω is the circular frequency, ρ_{11}^*, ρ_{22}^*, and ρ_{12}^* are equivalent masses to be explained later, $R = hE_2$, $Q = (1-h)E_2$, where h is the porosity and E_2 is the bulk modulus of the elasticity of the fluid in the pores, $P = A + 2N$, $A = \nu E_1/(1+\nu)(1-2\nu)$ is the first *Lamé constant*, $N = E_1/2(1+\nu)$ is the shear modulus of the solid phase of the porous material, E_1 and ν being the *in vacuo* Young's modulus and Poisson's ratio of the bulk solid phase.[1,15] Assuming that pores are cylindrical, an expression for E_2[5,6,14] has been defined as

$$E_2 = E_0\{1 + [2(\gamma - 1)/N_{\mathrm{Pr}}^{1/2} \lambda_c \sqrt{-j}] T_c[N_{\mathrm{Pr}}^{1/2} \lambda_c \sqrt{-j}]\}^{-1} \tag{5}$$

where $E_0 = \rho_0 c^2$, γ is the ratio of specific heats, c is the speed of sound in the fluid, $j = \sqrt{-1}$, and N_{Pr} represents the Prandtl number.

Among equivalent masses, ρ_{11}^* can be considered as the effective bulk mass[1] of the elastic part, ρ_{22}^* can be considered as the effective bulk mass[1] of the fluid part, and ρ_{12}^* can be considered as the coupling mass[1] between the frame and fluid, which are obtained as

$$\rho_{11}^* = \rho_{11} + b/j\omega, \quad \rho_{22}^* = \rho_{22} + b/j\omega, \quad \rho_{12}^* = \rho_{12} - b/j\omega, \tag{6}$$

where $\rho_{11} = \rho_1 + \rho_a$, $\rho_{22} = \rho_2 + \rho_a$, $\rho_{12} = -\rho_a$, and ρ_1 is the bulk density of the solid phase, $\rho_2 = h\rho_0$ is the bulk density of the fluid phase, ρ_0 is the density of interstitial fluid, and ρ_a is the coupling mass between the solid and fluid phases, which is

$$\rho_a = \rho_2(\varepsilon' - 1), \tag{7}$$

where ε' is a geometric factor that can be obtained from the tortuosity and porosity of the material.[2,7] b in Eq. (6) is a viscous coupling factor,[15,16] and is obtained from

$$b = j\omega\varepsilon'\rho_2(\rho_c^*/\rho_0 - 1), \tag{8}$$

where, for a cylindrical pore,[5,14]

$$\rho_c^* = \rho_0\{1 - (2/\lambda_c\sqrt{-j}) T_c[\lambda_c\sqrt{-j}]\}^{-1}. \tag{9}$$

λ_c in Eq. (9) is given as

$$\lambda_c^2 = 8\omega\rho_0\varepsilon'/h\sigma, \tag{10}$$

where σ is the steady-state, macroscopic flow resistivity. T_c in Eq. (9) is defined as

$$T_c[\lambda_c\sqrt{-j}] = J_1[\lambda_c\sqrt{-j}]/J_0[\lambda_c\sqrt{-j}], \tag{11}$$

where J_0 and J_1 are Bessel functions of first kind, zero and first order, respectively.

In vacuo Young's modulus E_1 is represented as $E_1 = E_m(1 + j\eta)$, where E_m is the static Young's modulus and η is the loss factor[24] to consider the material damping in the solid phase. Therefore, ρ_1, E_m, η, υ, σ, ε', h, and ρ_o are the set of macroscopic parameters that can be defined as the vibroacoustic characteristics of an elastic porous material.[5,6,14]

Because Eq. (1), which represents the propagation of dilatational waves in the solid phase with the influence of the fluid coupling terms, is a fourth-order equation, two dilatational wave types exist in the solid, which have wave numbers of

$$k_1^2 = \frac{(A_1 + \sqrt{A_1^2 - 4A_2})}{2}, \quad k_2^2 = \frac{(A_1 - \sqrt{A_1^2 - 4A_2})}{2}. \tag{12}$$

The wave related to k_1 is referred to as the airborne wave,[1,2] which describes the wave transmitted predominantly through the interstitial fluid, and the wave related to k_2 is referred to as the frame wave,[1,2] which describes the wave transmitted predominantly by the elastic structure. From Eq. (2), the wave number of the rotational wave is obtained as

$$k_t^2 = \frac{\omega^2}{N} \times [\rho_{11}^* - (\rho_{12}^*)^2/\rho_{22}^*]. \tag{13}$$

If we consider a layer of porous material with an infinite extent subjected to a plane wave input of the potential expressed as

$$\phi_i = e^{-j(k_x + k_y y)}, \tag{14}$$

where $k = \omega/c$, $k_x = k \times \sin\theta$, $k_y = k \times \cos\theta$, and θ is the angle of incidence (i.e., the angle from the normal to the surface). Because three types of wave exist in the porous material, six traveling (three forward and three backward) waves, which all have the same trace wave number, are induced by an oblique incident wave in a finite depth layer of porous material. x and y components of the displacements and stresses of the solid and fluid phases were derived as

follows by Bolton *et al.*[2] The displacements in the solid phase are

$$u_x = jk_x e^{-jk_x x}\left[\frac{C_1}{k_1^2}e^{-jk_{1y}y} + \frac{C_2}{k_1^2}e^{jk_{1y}y} + \frac{C_3}{k_2^2}e^{-jk_{2y}y} + \frac{C_4}{k_2^2}e^{jk_{2y}y}\right] - j\frac{k_{ty}}{k_t^2}e^{-jk_x x}[C_5 e^{-jk_{ty}y} - C_6 e^{jk_{1y}y}], \tag{15}$$

$$u_y = je^{-jk_x x}\left[\frac{k_{1y}}{k_1^2}C_1 e^{-jk_{1y}y} - \frac{k_{1y}}{k_1^2}C_2 e^{jk_{1y}y} + \frac{k_{2y}}{k_2^2}C_3 e^{-jk_{2y}y} - \frac{k_{2y}}{k_2^2}C_4 e^{jk_{2y}y}\right] + j\frac{k_x}{k_t^2}e^{-jk_x x}[C_5 e^{-jk_{ty}y} + C_6 e^{jk_{ty}y}]. \tag{16}$$

The displacements in the fluid phase are

$$U_x = jk_x e^{-jk_x x}\left[b_1\frac{C_1}{k_1^2}e^{-jk_{1y}y} + b_1\frac{C_2}{k_1^2}e^{jk_{1y}y} + b_2\frac{C_3}{k_2^2}e^{-jk_{2y}y} + b_2\frac{C_4}{k_2^2}e^{jk_{2y}y}\right] - jg\frac{k_{ty}}{k_t^2}e^{-jk_x x}[C_5 e^{-jk_{ty}y} - C_6 e^{jk_{ty}y}], \tag{17}$$

$$U_y = je^{-jk_x x}\left[b_1\frac{k_{1y}}{k_1^2}C_1 e^{-jk_{1y}y} - b_1\frac{k_{1y}}{k_1^2}C_2 e^{jk_{1y}y} + b_2\frac{k_{2y}}{k_2^2}C_3 e^{-jk_{2y}y} - b_2\frac{k_{2y}}{k_2^2}C_4 e^{jk_{2y}y}\right] + jg\frac{k_x}{k_t^2}e^{-jk_x x}[C_5 e^{-jk_{ty}y} + C_6 e^{jk_{ty}y}], \tag{18}$$

where in Eq. (18), b_1 and b_2 are defined[2] as

$$b_1 = \frac{(\rho_{11}^* - \rho_{12}^* Q)}{(\rho_{22}^* Q - \rho_{12}^* R)} - \frac{(PR - Q^2)}{[\omega^2(\rho_{22}^* Q - \rho_{12}^* R)]}k_1^2,$$

$$b_2 = \frac{(\rho_{11}^* R - \rho_{12}^* Q)}{(\rho_{22}^* Q - \rho_{12}^* R)} - \frac{(PR - Q^2)}{[\omega^2(\rho_{22}^* Q - \rho_{12}^* R)]}k_2^2.$$

The stresses in the solid phase are

$$\sigma_y = e^{-jk_x x}\left[\left(2N\frac{k_{1y}^2}{k_1^2} + A + b_1 Q\right)C_1 e^{-jk_{1y}y} + \left(2N\frac{k_{1y}^2}{k_1^2} + A + b_1 Q\right)C_2 e^{jk_{1y}y} + \left(2N\frac{k_{2y}^2}{k_2^2} + A + b_2 Q\right) \times C_3 e^{-jk_{2y}y} + \left(2N\frac{k_{2y}^2}{k_2^2} + A + b_2 Q\right)C_4 e^{jk_{2y}y} + 2N\frac{k_x k_{ty}}{k_t^2}(C_5 e^{-jk_{ty}y} - C_6 e^{jk_{ty}y})\right], \tag{19}$$

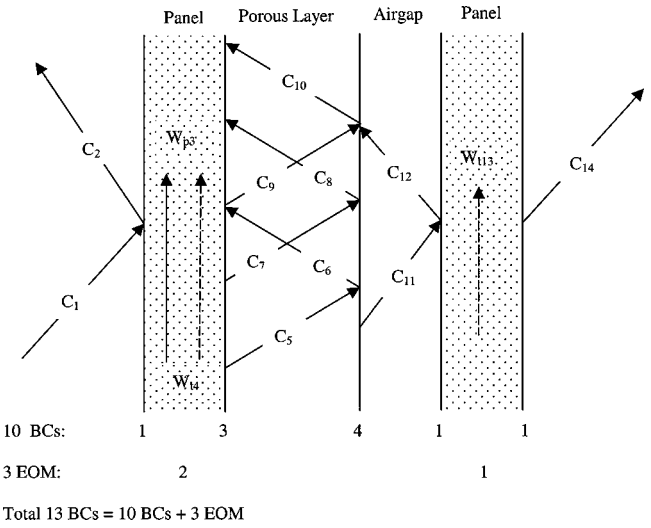


FIG. 1. Illustration of wave propagation in the B–U configuration.

$$\tau_{xy} = e^{-jk_x x}N\left[\frac{2k_x k_{1y}}{k_1^2}(C_1 e^{-jk_{1y}y} - C_2 e^{jk_{1y}y}) + \frac{2k_x k_{2y}}{k_2^2}(C_3 e^{-jk_{2y}y} - C_4 e^{jk_{2y}y}) + \frac{(k_x^2 - k_{ty}^2)}{k_t^2}(C_5 e^{-jk_{ty}y} + C_6 e^{jk_{ty}y})\right]. \tag{20}$$

The stresses in the fluid phase are

$$s = e^{-jk_x x}[(Q + b_1 R)C_1 e^{-jk_{1y}y} + (Q + b_1 R)C_2 e^{jk_{1y}y} + (Q + b_2 R)C_3 e^{-jk_{2y}y} + (Q + b_2 R)C_4 e^{jk_{2y}y}], \tag{21}$$

where σ_y can be considered as the force per unit material area acting in the y direction on the solid phase, s can be considered as the force per unit material area acting on the fluid component of the porous material, and τ_{xy} being the shear force acting on the solid phase.

The six constants $C_1 - C_6$ in Eqs. (15)–(21) have to be determined by applying boundary conditions (BCs), which therefore have to be represented in displacements and stresses. Bolton *et al.*[2,6,20,21] considered three types of boundary conditions associated with typical constructions of lined double panels. A bonded–bonded (B–B) construction is when the porous layer is bonded to both plates, a bonded-unbonded (B–U) construction is when one side of the porous layer is bonded and the other side is spaced from the plate, the unbonded–unbonded (U–U) construction is when both sides of the layer are spaced. Taking the BU shell as an example, traveling waves in the system and number of available boundary conditions at each boundary are shown in Fig. 1, from which the BB and UU shells can be deduced.

At the open surface of a porous layer (boundary between the porous layer and airgap in Fig. 1), four BCs exist, which are

(i) $-h \times P = s$; (ii) $-(1-h) \times P = \sigma_y$;

(iii) $\nu_y = j\omega(1-h)u_y + j\omega h U_y$; (iv) $\tau_{xy} = 0$;

where P is the pressure in the exterior acoustic field at the interface and v_y is the normal component of the acoustic particle velocity in the exterior medium at the interface.

When the elastic porous material is bonded directly to an elastic panel, there exist six BCs. Letting the panel transverse displacement be $w_t(x,t)=W_t(x)e^{j\omega t}$ and the centerline, in-plane, panel displacement be $w_p(x,t)=W_p(x)e^{j\omega t}$, those four BCs are

$$\text{(i)}\ \ v_y=j\omega W_t;\quad \text{(ii)}\ \ u_y=W_t;$$

$$\text{(iii)}\ \ U_y=W_t;\quad \text{(iv)}\ \ u_x=W_p(-/+)\frac{h_p}{2}\frac{dW_t}{dx}.$$

Two equations of motion become

$$\text{(v)}\ \ (+/-)\tau_{yx}=(D_p k_x^2-\omega^2 m_s)W_p;$$

$$\text{(vi)}\ \ (+/-)P(-/+)q_p-jk_x\frac{h_p}{2}\tau_{xy}=(Dk_x^4-\omega^2 m_s)W_t,$$

where h_p is the panel thickness, D is the bending stiffness per unit width, D_p is the membrane stiffness per unit width, m_s is the panel mass per unit area, and q_p is the normal force per unit panel area exerted on the panel by the elastic porous material ($q_p=-\sigma_y-s$). In BCs[2] (iv)–(vi), the first signs are appropriate when the porous material is attached to the positive y-facing surface of the panel, and the second signs when the porous material is attached to the negative y-facing surface.

In an elastic panel that is not in contact with the porous core on its either side, there are three BCs. Two of the BCs are

$$\text{(i)}\ \ v_{1y}=j\omega W_t;\ \ \text{(ii)}\ \ v_{2y}=j\omega W_t.$$

The third one is from the equation of motion:

$$\text{(iii)}\ \ P_1-P_2=(Dk_x^4-\omega^2 m_s)W_t,$$

where $P_{1,2}$ are the acoustic pressures in the air spaces on the two sides of the panel and $v_{1,2y}$ are the normal acoustic particle velocities at the two faces of the panels.

Therefore a total of 14 waves exist in the B–U panel and 13 BCs are available, as illustrated in Fig. 1, which permits us to calculate the pressure transmission coefficient. The transmission coefficient $\tau(\theta)$ is the ratio of the amplitudes of the incident and transmitted waves, for example, C_{14}/C_1 for the B–U panel, which is obviously a function of the incidence angle θ. To consider random incidence cases, the transmission coefficient $\tau(\theta)$ can be averaged according to the Paris formula[25]

$$\bar{\tau}=2\int_0^{\theta_m}\tau(\theta)\sin\theta\cos\theta\,d\theta \tag{22}$$

where θ_m is the maximum incident angle, which varies between 70° and 85° (72° in this paper), which is suggested by Mulholland *et al.*[26] Integration of Eq. (22) is conducted numerically by the Simpson's rule. Finally, the average TL is obtained as

$$\mathrm{TL}_{\mathrm{avg}}=10\log\frac{1}{\bar{\tau}}. \tag{23}$$

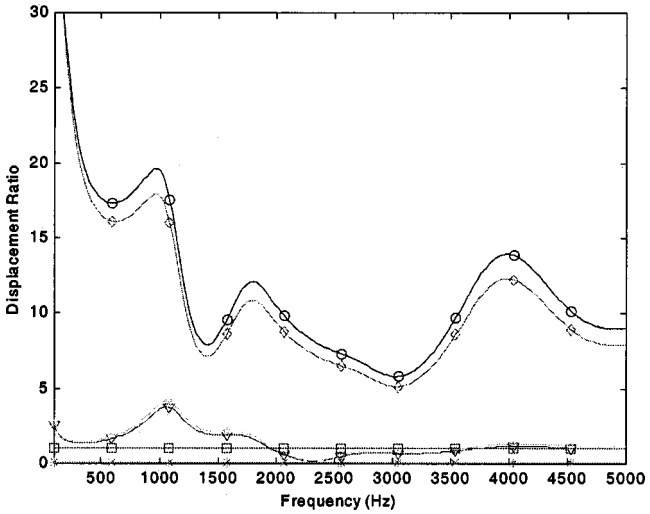


FIG. 2. Frame and shear wave contributions to the fluid and solid displacements in the y direction for the foam-lined double-panel B–B configuration (normalized by the strength of the airborne wave): □, $|b_1(k_{1y}/k_1^2)C_1|/|b_1(k_{1y}/k_1^2)C_1|$; ○, $b_2(k_{2y}/k_2^2)C_3|/|b_1(k_{1y}/k_1^2)C_1|$; ▽, $|g(k_x/k_1^2)C_5|/|b_1(k_{1y}/k_1^2)C_1|$, ∗, $|(k_{1y}/k_1^2)C_1|/|b_1(k_{1y}/k_1^2)C_1|$; ☆, $|(k_{2y}/k_2^2)C_3|/|b_1(k_{1y}/k_1^2)C_1|$; ◇, $|(k_x/k_1^2)C_5/|b_1(k_{1y}/k_1^2)C_1|$.

Bolton *et al.*[2] calculated TLs for three types of double panels and compared them with the measured TLs for the structures comprised of square aluminum sheets, 1.2 m on a side, of 1.27 and 0.762 mm thickness, and a 27 mm thick, partially reticulated, polyurethane foam slab. Values of other parameters can be found from Ref. 2. The solution process was reconstructed by the authors to be used for developing the approximate method and confirmed to provide the same results as in the original work. The TLs from the full method shown in Figs. 9–12 were calculated by the authors' reconstructed formulation, while the measured curve was taken from Ref. 2.

B. Development of the approximate analysis procedure

The approximate solution method is based on a quite simple idea, keeping only the most significant component among the three waves in the porous material. Because the most significant wave can be found only if the three waves are calculated, the simplified method has to be applied in two steps. Suppose the problem to solve sound transmission through a double walled cylindrical shell with a fiber core, which will be discussed as an example later. At first, a flat double panel with infinite extents with the same cross sectional construction is solved using the full method. Then, only the strongest wave number is chosen from the result and the material is modeled using the wave number and its corresponding equivalent density. Thus, the material is essentially modeled as an equivalent fluid.

Figure 2 compares the amplitudes of the six displacement components (three waves in the solid and three waves in the fluid), normalized by the amplitude of the airborne wave in the fluid phase for the B–B double panel. These ratios were obtained averaged over the incident angle to take into account the random incidence. It makes more sense to compare the significances of the waves in terms of the energy associated with the waves, which can be represented as follows

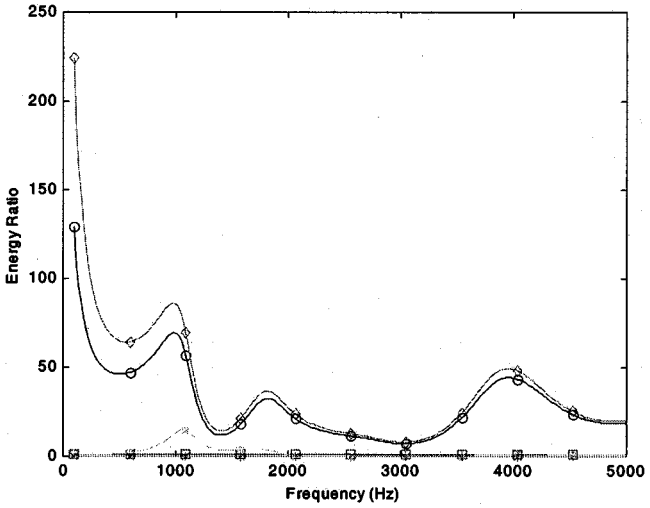


FIG. 3. Frame and shear wave contributions to the fluid and solid strain energies in the y direction for the foam-lined double-panel B–B configuration (normalized by the energy of the airborne wave): □, E_{1f}/E_{1f}; ○, E_{3f}/E_{1f}; *, E_{1s}/E_{1f}; ◇, E_{3s}/E_{1f}; ☆, E_{5s}/E_{1f}.

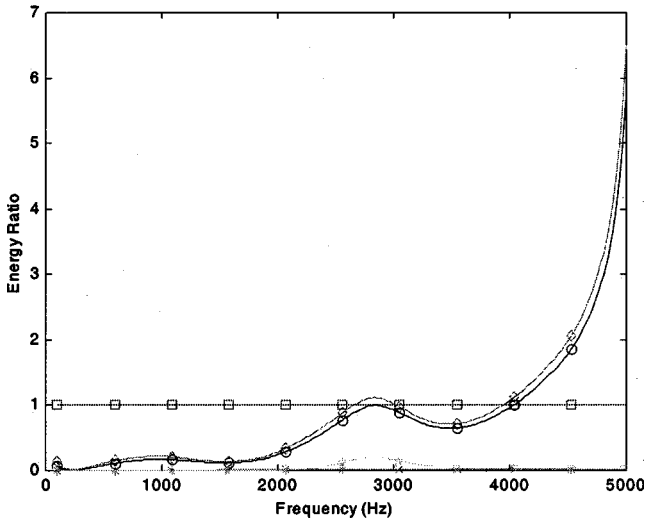


FIG. 4. Frame and shear wave contributions to the fluid and solid strain energies in the y direction for the foam-lined double-panel B–U configuration (normalized by the energy of the airborne wave): □, E_{1f}/E_{1f}; ○, E_{3f}/E_{1f}; *, E_{1s}/E_{1f}; ◇, E_{3s}/E_{1f}; ☆, E_{5s}/E_{1f}.

The airborne wave:

$$E_{1f}=\frac{1}{2}\left[h\cdot\left|(Q+b_1R)\cdot b_1\frac{k_{1y}^2}{k_1^2}C_1^2\right|\right]; \tag{24}$$

the frame wave:

$$E_{3f}=\frac{1}{2}\left[h\cdot\left|(Q+b_2R)\cdot b_2\frac{k_{2y}^2}{k_2^2}C_3^2\right|\right], \tag{25}$$

where the subscript f represents the fluid phase. Energy associated with the rotational wave (i.e., the shear wave) is not defined, since the rotational wave does not make a contribution to the fluid stress.

Similarly, the strain energy associated with the displacement in the solid phase can be defined for each wave component. For the airborne wave,

$$E_{1s}=\frac{1}{2}\left[(1-h)\cdot\left|\left(2N\frac{k_{1y}^2}{k_1^2}+A+b_1Q\right)\cdot\frac{k_{1y}^2}{k_1^2}C_1^2\right|\right]. \tag{26}$$

For the frame wave,

$$E_{3s}=\frac{1}{2}\left[(1-h)\cdot\left|\left(2N\frac{k_{2y}^2}{k_2^2}+A+b_2Q\right)\cdot\frac{k_{2y}^2}{k_2^2}C_3^2\right|\right]. \tag{27}$$

For the rotational wave,

$$E_{5s}=\frac{1}{2}\left[(1-h)\cdot\left|2N\left(\frac{k_{2y}^2}{k_2^2}\right)^2C_5^2\right|\right], \tag{28}$$

where the subscript s represents the solid phase.

Figures 3–5 show the ratios of the energy carried by the frame wave and the rotational wave to the airborne wave in the fluid and solid phases: i.e., E_{1f}/E_{1f}, E_{3f}/E_{1f}, E_{1s}/E_{1f}, E_{3s}/E_{1f}, E_{5s}/E_{1f} for the B–B, B–U, and U–U configurations. From the figures, it is clear that the frame wave (i.e., k_2) is the most significant component in the entire frequency range in the B–B configuration. For the B–U configuration, the airborne wave (i.e., k_1) is found to be the strongest component in the low-frequency range, while the frame wave (i.e., k_2) is the strongest component in the high-frequency range. In the U–U configuration, the airborne wave (i.e., k_1) is dominant over the entire frequency range. In all three cases, the shear wave (rotational wave) is the least contributing component. It is noted that the strongest wave component and the frequency ranges related depend on the construction of the double panel, therefore the analysis by the full method has to be conducted for each new problem.

Using only the strongest wave, the porous material is now modeled as an equivalent fluid, whose speed of sound and density are determined as those corresponding to the wave retained. For example, in the B–B panel of the given dimension, the wave number k_2 in Eq. (12) and the bulk density of the solid phase of the porous material (i.e., ρ_{11}^*) may be used. The wave number k_1 in Eq. (12) and the bulk density of the fluid phase (i.e., ρ_{22}^*) may be used if the airborne wave is the strongest wave as in the U–U case. The

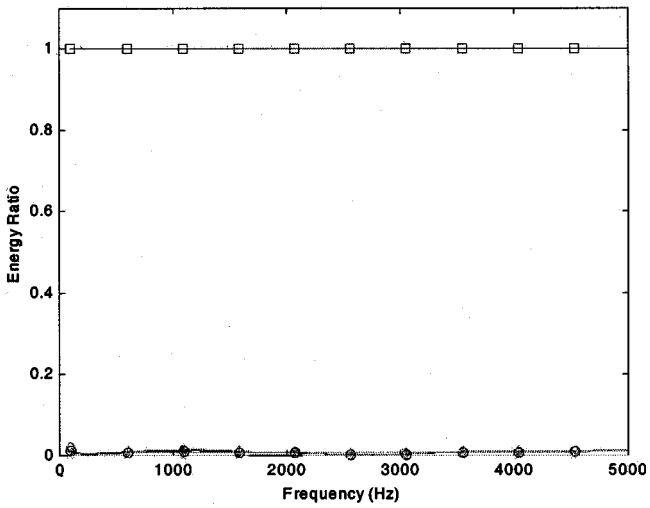


FIG. 5. Frame and shear wave contributions to the fluid and solid strain energies in the y direction for the foam-lined double-panel U–U configuration (normalized by the energy of the airborne wave): □, E_{1f}/E_{1f}; ○, E_{3f}/E_{1f}; *, E_{1s}/E_{1f}; ◇, E_{3s}/E_{1f}; ☆, E_{5s}/E_{1f}.

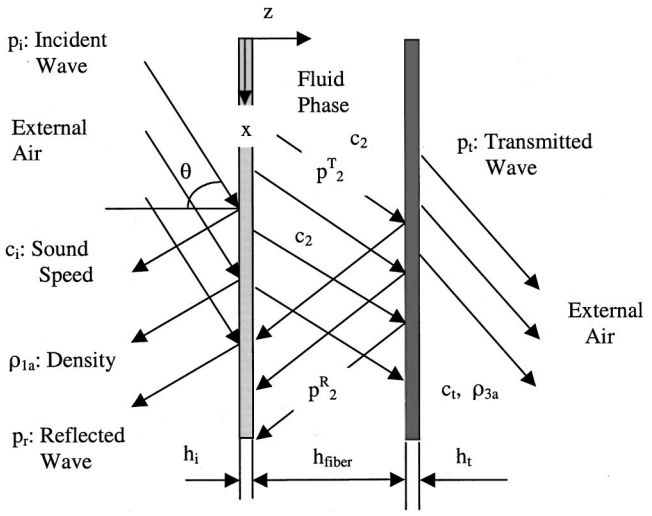


FIG. 6. Simplified B–B model of the double panel lined with a porous layer.

complex speed of sound is calculated from the wave number and the density, if necessary.

To explain how this simplified method can be used in a practice, consider a curved double panel lined with a porous core with a B–U construction. The modeling procedure will be the following.

(1) Build a flat double panel model of infinite extent with a B–U core, keeping all the other parameters (thickness of the shell, airgap, and the porous layer, elastic properties of the shell etc.) the same.

(2) Solve the problem by the full method to obtain the ratios of the energy strengths of the three waves. Choose the strongest wave and use its wave number and the corresponding equivalent mass to describe the porous material.

(3) Build the double shell model treating the porous material as an equivalent fluid. The system model will become a double shell enclosing two layers of fluid, one representing the equivalent fluid to the porous material and the other representing the air in the gap.

After the simplification as stated above, the problem can be solved in a straightforward way using existing analysis tools. For example, any commercial FEA software can be used if it has the capability to analyze a regular acoustics element. Obviously the proposed approximate analysis method compromises accuracy for simplicity.

III. ANALYSIS OF SOUND TRANSMISSION THROUGH DOUBLE PANELS USING THE SIMPLIFIED METHOD

A. Formulation of the problem

1. Double panel with bonded–bonded (B–B) layer

Figure 6 shows the simplified model of a double panel lined with a porous layer in the B–B configuration. The mass density and sound speed of the acoustic media in the incident and transmitted sides are $\{\rho_{1a},c_i\}$ and $\{\rho_{3a},c_t\}$. An oblique wave p_i is incident with an angle θ (measured from the normal axes to the beam) from one side, which induces the traveling transverse waves in the plates w_1^0 and w_2^0, the reflected acoustic wave p_r, transmitted wave p_t in the external acoustic spaces, and the reflected acoustic wave p_2^R, and transmitted wave p_2^T in the porous layer.

In the air incident side, the wave equation[27] becomes

$$c_i\nabla^2(p_i+p_r)+\frac{\partial^2(p_i+p_r)}{\partial t^2}=0, \tag{29}$$

where ∇^2 is the Laplacian operator in the rectangular coordinate and p_r is the reflected wave. The wave equation in the porous material layer becomes

$$c_2\nabla^2(p_2^T+p_2^R)+\frac{\partial^2(p_2^T+p_2^R)}{\partial t^2}=0. \tag{30}$$

Note that c_2 is the speed of sound in the equivalent fluid obtained by taking the strongest wave in the porous material obtained by the full theory as explained in Sec. II.

In the air in the transmitted side, the wave equation becomes

$$c_t\nabla^2 p_t+\frac{\partial^2 p_t}{\partial t^2}=0. \tag{31}$$

Equation of motion of the plate[28] of the incident side is given as

$$-D_i\frac{\partial w_1^0}{\partial x^4}+[(p_i+p_r)-(p_2^T+p_2^R)]=\rho_i h_i\ddot{w}_1^0. \tag{32}$$

Equation of motion of the plate on the transmitted side is given as

$$-D_t\frac{\partial w_2^0}{\partial x^4}+[(p_2^T+p_2^R)-p_t]=\rho_t h_t\ddot{w}_2^0, \tag{33}$$

where w_1^0, w_2^0 are the transverse displacements of the plates, D_i and D_t are the flexural rigidity of the plates, ρ_i and ρ_t are the material density of the plates, h_i and h_t are the thickness of the plates, and the dot notation indicates the derivative with respect to time.

The BC between the air and the plate at the incident side is given as

$$\frac{\partial(p_i+p_r)}{\partial z}=-\rho_{1a}\frac{\partial^2 w_1^0}{\partial t^2},\quad \text{at } z=0. \tag{34}$$

Between the plate and the inside fluid, which is idealizing the porous material,

$$\frac{\partial(p_2^T+p_2^R)}{\partial z}=-\rho_{11}^*\frac{\partial^2 w_1^0}{\partial t^2},\quad \text{at } z=0, \tag{35}$$

where ρ_{11}^* is the bulk density of the solid phase of the porous material, which is selected because the dominant wave is the frame wave in this case as discussed. The BC between the internal fluid and the transmitted side plate is given as

$$\frac{\partial(p_2^T+p_2^R)}{\partial z}=-\rho_{11}^*\frac{\partial^2 w_2^0}{\partial t^2},\quad \text{at } z=h_{\text{fiber}}. \tag{36}$$

Between the transmitted side plate and the air,

$$\frac{\partial p_t}{\partial z}=-\rho_{3a}\frac{\partial^2 w_2^0}{\partial t^2},\quad \text{at } z=h_{\text{fiber}}. \tag{37}$$

The harmonic incident wave[27,29] is expressed as

$$p_i(x,z,t)=p_0^{j(\omega t-k_{1x}x-k_{1z}z)}, \tag{38}$$

where p_0 is the effective amplitude of the incident wave; ω angular or rotational frequency; and

$$k_1=\frac{\omega}{c_1}, \quad k_{1x}=k_1\sin(\theta), \quad k_{1z}=k_1\cos(\theta). \tag{39}$$

In the porous layer, the frame wave number obtained in Eq. (12) is taken as the equivalent wave number k_2 because the frame wave is the dominant carrier. Therefore,

$$k_{2x}=k_{1x}, \quad k_{2z}=\sqrt{k_2^2-k_{2x}^2}. \tag{40}$$

In the transmitted side,

$$k_3=\frac{\omega}{c_3}, \quad k_{3x}=k_{1x}, \quad k_{3z}=\sqrt{k_3^2-k_{3x}^2}. \tag{41}$$

Thus, p_r, p_2^T, p_2^R, p_t, w_1^0, and w_2^0 can be expressed as follows:

$$p_r(x,z,t)=p_{ic}^R e^{j(\omega t-k_{1x}x+k_{1z}z)}, \tag{42}$$

$$p_2^R(x,z,t)=p_{2c}^R e^{j(\omega t-k_{1x}x+k_{2z}z)}, \tag{43}$$

$$p_2^T(x,z,t)=p_{2c}^T e^{j(\omega t-k_{1x}x-k_{2z}z)}, \tag{44}$$

$$p_t(x,z,t)=p_{tc}^T e^{j(\omega t-k_{1x}x-k_{3z}z)}, \tag{45}$$

$$w_1^0(z,t)=w_{1c}^0 e^{j(\omega t-k_{1x}x)}, \tag{46}$$

$$w_2^0(z,t)=w_{2c}^0 e^{j(\omega t-k_{1x}x)}. \tag{47}$$

By substituting Eqs. (38), (42)–(47) into Eqs. (32)–(33) and (34)–(37), which are two equations of motion of the beams and four BCs, six equations are obtained for six unknown complex amplitudes p_{ic}^R, p_{2c}^T, p_{2c}^R, p_{tc}^T, w_{1c}^0, and w_{2c}^0, which can be put into a matrix equation as follows:

$$\begin{bmatrix} A & B & C & 0 & D & 0 \\ 0 & F & G & H & 0 & I \\ J & 0 & 0 & 0 & K & 0 \\ 0 & M & N & 0 & O & 0 \\ 0 & P & Q & 0 & 0 & R \\ 0 & 0 & 0 & S & 0 & T \end{bmatrix} \begin{Bmatrix} p_{ic}^R \\ P_{2c}^T \\ p_{2c}^R \\ p_{tc}^T \\ w_{1c}^0 \\ w_{2c}^0 \end{Bmatrix} = \begin{Bmatrix} E \\ 0 \\ L \\ 0 \\ 0 \\ 0 \end{Bmatrix}, \tag{48}$$

where

$$A=1, \quad B=-1, \quad C=-1, \quad D=\rho_i h_i\omega^2-D_i k_{1x}^4,$$

$$E=-p_0, \quad F=e^{-jk_{2z}h_{\text{fiber}}}, \quad G=e^{jk_{2z}h_{\text{fiber}}},$$

$$H=-e^{-jk_{3z}h_{\text{fiber}}}, \quad I=\rho_t h_t\omega^2-D_t k_{1x}^4, \quad J=jk_{1z},$$

$$K=-\rho_{1a}\omega^2, \quad L=jk_{1z}p_0, \quad M=-jk_{2z},$$

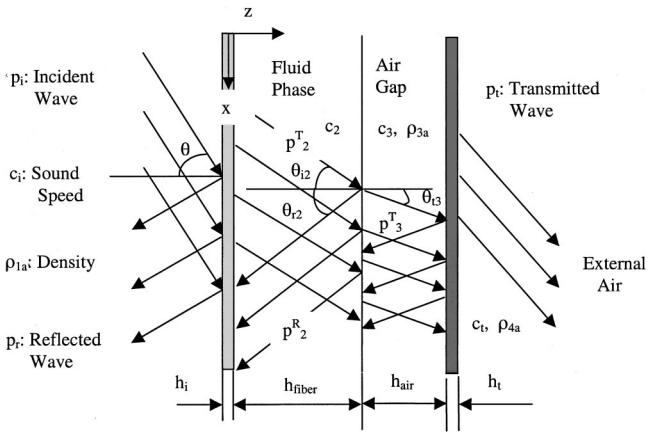


FIG. 7. Simplified B–U model of the double panel lined with a porous layer.

$$N=jk_{2z}, \quad O=-\rho_{11}^*\omega^2,$$

$$P=-jk_{2z}e^{-jk_{2z}h_{\text{fiber}}}, \quad Q=jk_{2z}e^{jk_{2z}h_{\text{fiber}}},$$

$$R=-\rho_{11}^*\omega^2, \quad S=jk_{3z}e^{-jk_{3z}h_{\text{fiber}}}, \quad T=\rho_{3a}\omega^2.$$

The six unknown coefficients p_{ic}^R, p_{2c}^T, p_{2c}^R, p_{tc}^T, w_{1c}^0, and w_{2c}^0 are obtained in terms of p_o by solving Eq. (48). These can be substituted back into Eqs. (42)–(47) to find the displacements of the shell and the acoustic pressures.

2. Bonded–unbonded (B–U) configuration

Figure 7 shows a double-panel structure lined with a porous layer in the B–U configuration. The mass density and sound speed of the acoustic media in the incident, transmitted sides, and airgap space are $\{\rho_{1a},c_i\}$, $\{\rho_{4a},c_t\}$, and $\{\rho_{3a},c_3\}$. The mass density of the fluid phase of the porous layer is defined as ρ_{22}^* if low-frequency sound is the major concern and ρ_{11}^* if high-frequency sound is the major concern, where ρ_{11}^* and ρ_{22}^* are defined in Eq. (6).

The procedure to derive governing equations is very similar to the BB case except that the interface conditions between the layers of the equivalent fluid and the internal air are necessary:

$$p_2^T+p_2^R=p_3^T+p_3^R, \quad \text{at } z=h_{\text{fiber}}, \tag{49}$$

$$u_2^T+u_2^R=u_3^T+u_3^R, \quad \text{at } z=h_{\text{fiber}}, \tag{50}$$

where u indicates the particle velocity in the acoustic space. Also for this case, the airborne wave or frame wave is the major energy carrier depending on the frequency of interest. In this particular case, the airborne wave has to be taken for the frequencies lower than 2.7 kHz (see Fig. 7), and the frame wave if higher than 2.7 kHz.

These eight equations can be put into a matrix form as follows:

$$\begin{bmatrix} A & B & C & 0 & 0 & 0 & D & 0 \\ 0 & 0 & 0 & F & G & H & 0 & I \\ J & 0 & 0 & 0 & 0 & 0 & K & 0 \\ 0 & M & N & 0 & 0 & 0 & O & 0 \\ 0 & P & Q & R & S & 0 & 0 & 0 \\ 0 & T & U & V & W & 0 & 0 & 0 \\ 0 & 0 & 0 & X & Y & 0 & 0 & Z \\ 0 & 0 & 0 & 0 & 0 & A1 & 0 & B1 \end{bmatrix} \begin{Bmatrix} p_{ic}^{R} \\ p_{2c}^{T} \\ p_{2c}^{R} \\ p_{3c}^{T} \\ p_{3c}^{R} \\ p_{tc}^{T} \\ w_{1c}^{0} \\ w_{2c}^{0} \end{Bmatrix} = \begin{Bmatrix} E \\ 0 \\ L \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix}, \tag{51}$$

where, for low-frequency sound,

$$A=1, \quad B=-1, \quad C=-1, \quad D=\rho_i h_i \omega^2 - D_i k_{1x}^4,$$

$$E=-p_0, \quad F=e^{-jk_{3z}(h_{\text{fiber}}+h_{\text{air}})},$$

$$G=e^{jk_{3z}(h_{\text{fiber}}+h_{\text{air}})}, \quad H=-e^{-jk_{4z}(h_{\text{fiber}}+h_{\text{air}})},$$

$$I=(\rho_t h_t \omega^2 - D_t k_{1x}^4), \quad J=jk_{1z}, \quad K=-\rho_{1a}\omega^2,$$

$$L=jk_{1z}p_0, \quad M=-jk_{2z} \quad N=jk_{2z}, \quad O=-\rho_{22}^{*}\omega^2,$$

$$P=e^{-jk_{2z}h_{\text{fiber}}}, \quad Q=e^{jk_{2z}h_{\text{fiber}}}, \quad R=-e^{-jk_{3z}h_{\text{fiber}}},$$

$$S=-e^{jk_{3z}h_{\text{fiber}}}, \quad T=\frac{\cos\theta_{i2}}{\rho_{22}^{*}c_2}e^{-jk_{2z}h_{\text{fiber}}},$$

$$U=-\frac{\cos\theta_{r2}}{\rho_{22}^{*}c_2}e^{jk_{2z}h_{\text{fiber}}}, \quad V=-\frac{\cos\theta_{t3}}{\rho_3 c_3}e^{-jk_{3z}h_{\text{fiber}}},$$

$$W=\frac{\cos\theta_{r3}}{\rho_3 c_3}e^{jk_{3z}h_{\text{fiber}}}, \quad X=-jk_{3z}e^{-jk_{3z}(h_{\text{fiber}}+h_{\text{air}})},$$

$$Y=jk_{3z}e^{jk_{3z}(h_{\text{fiber}}+h_{\text{air}})}, \quad Z=-\rho_{3a}\omega^2,$$

$$\text{A1}=jk_{4z}e^{-jk_{4z}(h_{\text{fiber}}+h_{\text{air}})}, \quad \text{B1}=\rho_{4a}\omega^2.$$

For high-frequency sound, the terms of O, T, and U should be replaced by

$$O=-\rho_{11}^{*}\omega^2, \quad T=\frac{\cos\theta_{i2}}{\rho_{11}^{*}c_2}e^{-jk_{2z}h_{\text{fiber}}},$$

$$U=\frac{\cos\theta_{r2}}{\rho_{11}^{*}c_2}e^{jk_{2z}h_{\text{fiber}}},$$

where θ_{i2}, θ_{r2}, and θ_{t3}, which are the angles of the incident, reflected, and transmitted waves and are related to one another as

$$\cos\theta_{i2}=\frac{k_{2z}}{k_2}, \quad \sin\theta_{i2}=\sin\theta_{r2},$$

$$\frac{\sin\theta_{i2}}{c_2}=\frac{\sin\theta_{t3}}{c_3}, \quad \sin\theta_{t3}=\sin\theta_{r3}. \tag{52}$$

3. Unbonded–unbonded (U–U) configuration

Figure 8 shows a double-panel structure lined with a porous layer in the U–U configuration. h_{air1} and h_{air2} are the depth of the airgaps as shown. The mass density and sound speed of the acoustic media in the incident and transmitted sides are $\{\rho_{1a}, c_i\}$ and $\{\rho_{5a}, c_t\}$. The mass density and sound speed of the airgap spaces in the incident side and transmitted side are $\{\rho_{2a}, c_2\}$ and $\{\rho_{4a}, c_4\}$. Because the major energy carrier of the U–U panel is the airborne wave, the wave number is taken from Eq. (12), and the effective mass density of the fluid phase of the porous layer ρ_{22}^{*}, which is given in Eq. (6). Governing equations can be derived by a very similar procedure as the B–U case, except that there is one more equation because of the additional airgap, which results in the following 10×10 matrix equation for ten unknown complex amplitudes p_{ic}^{R}, p_{2c}^{T}, p_{2c}^{R}, p_{3c}^{T}, p_{3c}^{R}, p_{4c}^{T}, p_{4c}^{R}, p_{tc}^{T}, w_{1c}^{0}, and w_{2c}^{0}.

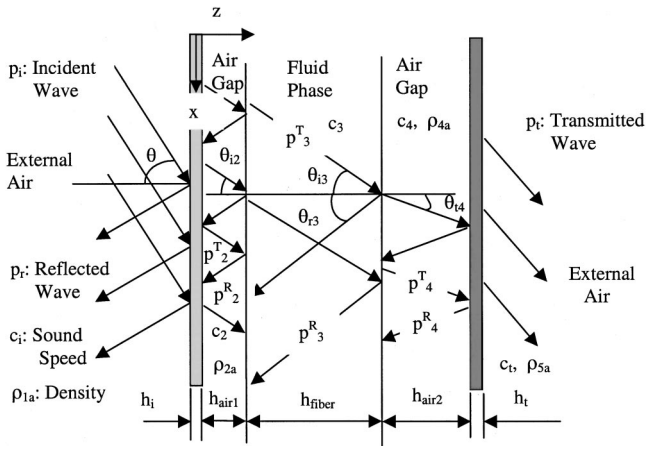


FIG. 8. Simplified U–U model of the double panel lined with a porous layer.

$$\begin{bmatrix} A & B & C & 0 & 0 & 0 & 0 & 0 & D & 0 \\ 0 & 0 & 0 & 0 & 0 & F & G & H & 0 & I \\ J & 0 & 0 & 0 & 0 & 0 & 0 & 0 & K & 0 \\ 0 & M & N & 0 & 0 & 0 & 0 & 0 & O & 0 \\ 0 & P & Q & R & S & 0 & 0 & 0 & 0 & 0 \\ 0 & T & U & V & W & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & X & Y & Z & \text{A1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \text{B1} & \text{C1} & \text{D1} & \text{E1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \text{F1} & \text{G1} & 0 & 0 & \text{H1} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \text{I1} & 0 & \text{J1} \end{bmatrix} \begin{Bmatrix} p_1^R \\ p_2^T \\ p_2^R \\ p_3^T \\ p_3^R \\ p_4^T \\ p_4^R \\ p_5^T \\ w_1^0 \\ w_2^0 \end{Bmatrix} = \begin{Bmatrix} E \\ 0 \\ L \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix}, \tag{53}$$

where

$$A=1, \quad B=-1, \quad C=-1, \quad D=\rho_i h_i \omega^2 - D_i k_{1x}^4,$$

$$E=-p_0, \quad F=e^{-jk_{4z}(h_{\text{air1}}+h_{\text{fiber}}+h_{\text{air2}})},$$

$$G=e^{jk_{4z}(h_{\text{air}}+h_{\text{fiber}}+h_{\text{air2}})},\quad H=-e^{jk_{5z}(h_{\text{air1}}+h_{\text{fiber}}+h_{\text{air2}})},$$

$$I=(\rho_t h_t\omega^2-D_t k_{1x}^4),\quad J=jk_{1z},\quad K=-\rho_{1a}\omega^2,$$

$$L=jk_{1z}p_0,\quad M=-jk_{2z},\quad N=jk_{2z},$$

$$O=-\rho_{2a}\omega^2,\quad P=e^{-jk_{2z}h_{\text{air1}}},\quad Q=e^{jk_{2z}h_{\text{air1}}},$$

$$R=-e^{-jk_{3z}h_{\text{air1}}},\quad S=-e^{jk_{3z}h_{\text{air1}}},$$

$$T=\frac{\cos\theta_{i2}}{\rho_2 c_2}e^{-jk_{2z}h_{\text{air1}r}},\quad U=-\frac{\cos\theta_{r2}}{\rho_2 c_2}e^{jk_{2z}h_{\text{air1}}},$$

$$V=-\frac{\cos\theta_{t3}}{\rho_{22}^* c_3}e^{-jk_{3z}h_{\text{air1}}},\quad W=\frac{\cos\theta_{r3}}{\rho_{22}^* c_3}e^{jk_{3z}h_{\text{air1}}},$$

$$X=e^{-jk_{3z}(h_{\text{air1}}+h_{\text{fiber}})},\quad Y=e^{jk_{3z}(h_{\text{air1}}+h_{\text{fiber}})},$$

$$Z=-e^{-jk_{4z}(h_{\text{air1}}+h_{\text{fiber}})},\quad \text{A1}=-e^{jk_{4z}(h_{\text{air1}}+h_{\text{fiber}})},$$

$$\text{B1}=\frac{\cos\theta_{t3}}{\rho_{22}^* c_3}e^{-jk_{3z}(h_{\text{air1}}+h_{\text{fiber}})},$$

$$\text{C1}=-\frac{\cos\theta_{r3}}{\rho_{22}^* c_3}e^{jk_{3z}(h_{\text{air1}}+h_{\text{fiber}})},$$

$$\text{D1}=-\frac{\cos\theta_{t4}}{\rho_{4a} c_4}e^{-jk_{4z}(h_{\text{air1}}+h_{\text{fiber}})},$$

$$\text{E1}=\frac{\cos\theta_{r4}}{\rho_{4a} c_4}e^{jk_{4z}(h_{\text{air1}}+h_{\text{fiber}})},$$

$$\text{F1}=-jk_{4z}e^{-jk_{4z}(h_{\text{air1}}+h_{\text{fiber}}+h_{\text{air2}})},$$

$$\text{G1}=jk_{4z}e^{jk_{4z}(h_{\text{air1}}+h_{\text{fiber}}+h_{\text{air2}})},$$

$$\text{H1}=-\rho_{4a}\omega^2,\quad \text{I1}=jk_{5z}e^{-jk_{5z}(h_{\text{air1}}+h_{\text{fiber}}+h_{\text{air2}})},$$

$$\text{J1}=\rho_{5a}\omega^2.$$

The relationships[27] among θ_{i2}, θ_{r2}, and θ_{t3} are found as

$$\cos\theta_{i2}=\frac{k_{2z}}{k_2},\quad \sin\theta_{i2}=\sin\theta_{r2},$$
$$\frac{\sin\theta_{i2}}{c_2}=\frac{\sin\theta_{t3}}{c_3},\quad \sin\theta_{t3}=\sin\theta_{r3}. \tag{54}$$

Similarly, the relationships[27] among θ_{t3}, θ_{r3}, and θ_{t4} are found as

$$\cos\theta_{t3}=\frac{k_{3z}}{k_3},\quad \sin\theta_{t3}=\sin\theta_{r3},$$
$$\frac{\sin\theta_{t3}}{c_3}=\frac{\sin\theta_{t4}}{c_4},\quad \sin\theta_{t4}=\sin\theta_{r4}. \tag{55}$$

B. Comparison to the results from the full theory

Figure 9 shows the TL calculated using the simplified theory compared to the TLs obtained from the full method and the measurement (taken from Ref. 2) for the B–B double panel. Averaging according to the diffusive field theory was taken [see Eq. (22)] to calculate the TLs in the random incidence condition, which was the measured condition. As it is shown, the TL curve obtained from the simplified model lies generally within 5 dB from the other TL curves, and the characteristics of the curves match with one another very well.

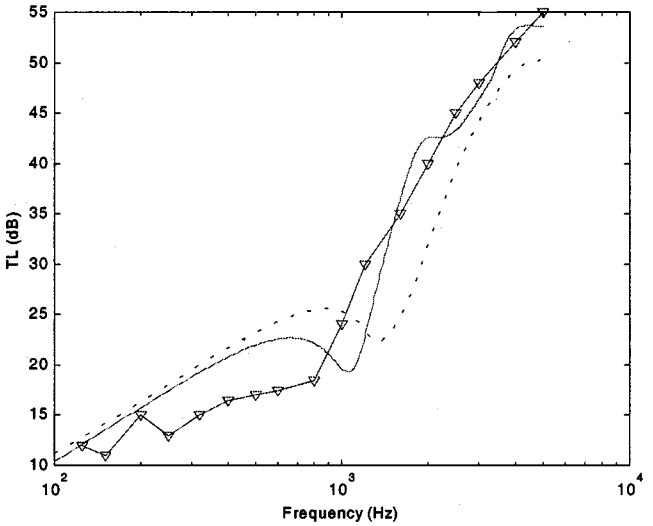


FIG. 9. A comparison of TLs of the double-panel B–B configuration. ▽, measured result; ---, prediction of the simplified model; —, prediction of the full model.

One possible argument may be whether the proposed method is really better than other approximate methods that have been used. For example, only the structural effect of the porous layer was considered in some studies, adding half of the mass of the porous layer to each panel and considering the continuous spring between the plates[30] to represent the stiffness of the layer. Figure 10 compares the TL obtained from such a model, the TL obtained from the simplified model proposed in this work, and the TL obtained from the full method for the same B–B double panel in one graph. From the comparison in Fig. 10, it is seen that the simplified model proposed in this work provides a much better approximation, which is expected because the model considers the characteristics of the porous material, even though it is in an approximate sense.

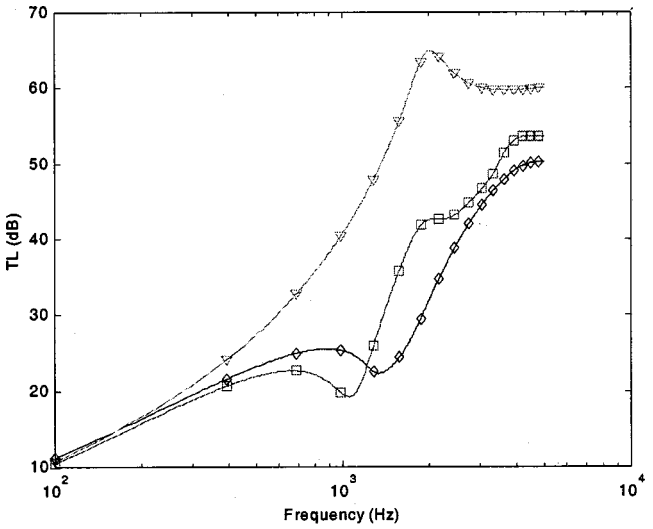


FIG. 10. A comparison of calculated TLs of the double-panel B–B configuration. □, the full model; ◇, the simplified model; ▽, the approximate model.

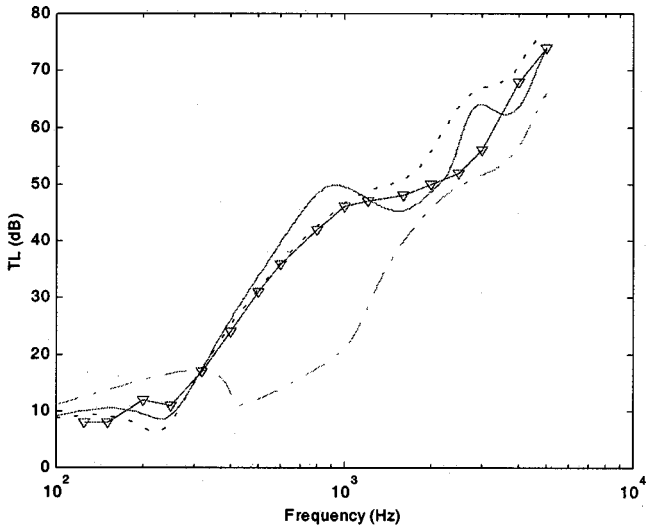


FIG. 11. A comparison of TLs of the double-panel B–U configuration. ▽, measured result; ---, prediction of the simplified model using the airborne wave; –·–, prediction of the simplified model using the frame wave; —, prediction of the full model.

The TL obtained from the simplified theory is again compared to those from the full theory and experiment for the B–U configuration in Fig. 11. In the figure, the TL using the airborne wave as the major energy carrier and the TL obtained by using the frame wave as the major energy carrier are compared with the TL curves obtained from the full method and experiment. It is seen that a good agreement is obtained in the low-frequency range for the former and a good agreement is obtained in the high-frequency range for the latter, as predicted in Sec. II. Thus, the two cases will have to be used depending on the frequency of interest.

Figure 12 shows the TLs for the U–U case, in which the airborne wave is used as the only energy carrier. It is seen that the simplified model provides a very good agreement with the full theory and measured results. This is expected because the airborne wave is stronger by far than the other two waves in the U–U panel.

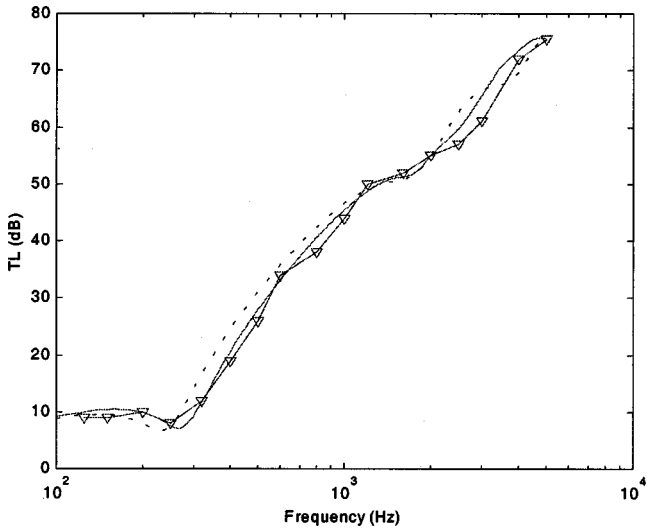


FIG. 12. A comparison of TLs of the double-panel U–U configuration. ▽, measured result; ---, prediction of the simplified model; —, prediction of the full model.

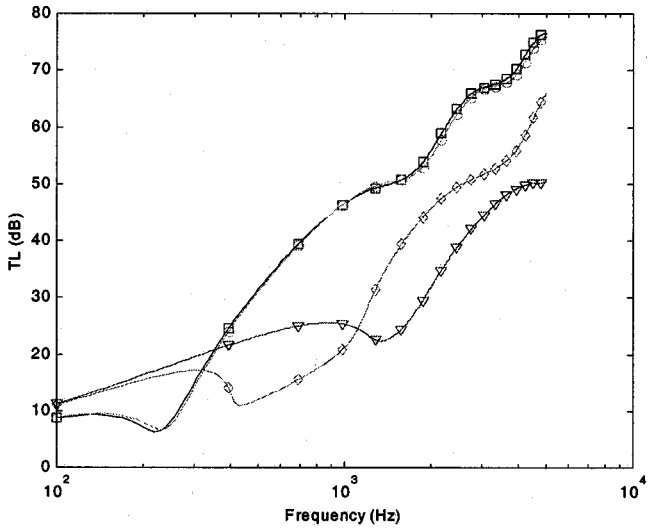


FIG. 13. Summary of the calculated TLs of three double-panel configurations. ▽, B–B; □, B–U (airborne wave); ◇, B–U (frame wave); ○, U–U.

IV. EXAMPLES: SOUND TRANSMISSION THROUGH DOUBLE WALLED CYLINDRICAL SHELLS WITH POROUS CORES

A double walled cylindrical shell whose walls sandwiching a layer of porous material is found in applications that require a high level of noise insulation such as the aircraft fuselage. The approximate analysis method developed in this work is applied to calculate the TLs through such structures. Analytical studies on the vibration or acoustic responses of cylindrical structures of various cross sections are found abundantly,[31–38] which typically approximated the problem as an infinitely long cylinder subjected to a plane wave incident with an angle γ from outside and the anechoic condition inside, as shown in Fig. 13. Three types of core constructions are considered, which are a bonded–bonded (B–B) shell, in which the layer is bonded to both shells, a bonded–unbonded (B–U) shell, in which the layer is bonded to the external shell but separated by an airgap from the internal shell, and a unbonded–unbonded (U–U) shell, in which airgaps exist on both sides of the layer. The Love's equation was used to describe the shell motions and structural–acoustic coupling effects were fully considered. The actual formulation procedure and governing equations can be found from the authors' previous works on the sound transmission through double walled cylinders of various sections.[39,40]

A. B–B shell

As explained, a preliminary calculation was conducted using the full theory for a flat double panel of infinite extent with the B–B core, which identified the frame wave as the wave component to represent the porous material as an equivalent fluid. $R_e = 0.2$ m, $R_i = 0.15$ m, which are the radii of the external and internal shells, $h_e = 2$ mm, $h_i = 3$ mm, are used for the calculation. Internal and external fluid media are air at 20 °C.

Figure 14 shows the TL of the B–B shell compared to the TL of the double walled shell without the porous material, which shows that adding the porous material to the double shell in a B–B configuration is not so effective to

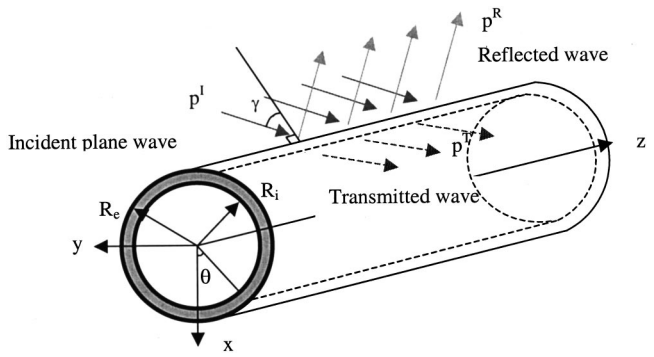


FIG. 14. Schematic diagram of the double shell subjected to a plane incident wave.

increase the TL, except in the frequency range lower than 250 Hz. It is seen that adding the porous layer even slightly decreases the TL in the frequency range between 250 and 3000 Hz. However, this may have been caused by the fact that the simplified model tends to underestimate the TL in the high-frequency range, as seen in Fig. 10. Considering the approximations involved, it can be concluded that applying the porous material to the double shell in a B–B construction, without leaving any airgap between the layers, would make only marginal improvement in the low-frequency range.

B. B–U shell

Figure 15 shows the TLs calculated for a BU shell. The same dimensions as the case of Fig. 14 were used, except that $R_c=0.169$ m and $h_c=30$ mm for the radius of the center of the core and thickness of the porous material, therefore an airgap of 17.5 mm between the internal shell and the porous layer. As in the flat panel case, the airborne wave is used in the low-frequency range, and the frame wave is used in the high-frequency range to represent the porous material as an equivalent fluid, with the transit frequency of about 1800 Hz below which the airborne wave is the main energy carrier,

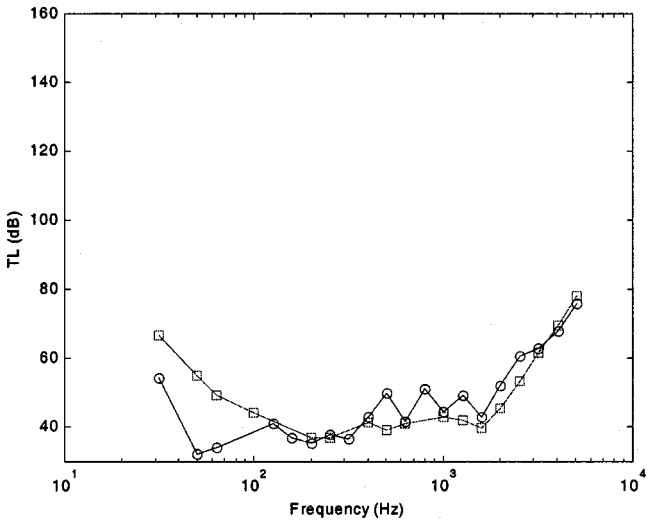


FIG. 15. A comparison of calculated TLs of cylindrical double walled shells ($R_e=0.2$ m, $R_i=0.15$ m, $h_e=2$ mm, $h_i=3$ mm, and $h_c=47.5$ mm). ○, no foam; □, B–B.

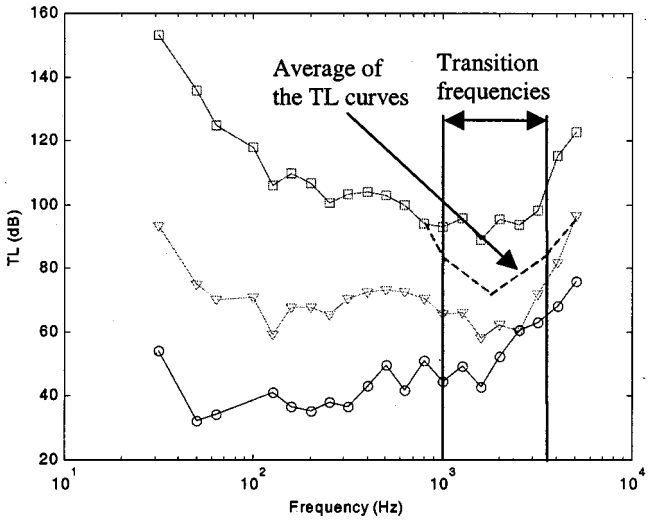


FIG. 16. Comparison of calculated TLs of cylindrical double walled shells ($R_e=0.2$ m, $R_i=0.15$ m, $R_c=0.169$ m, $h_e=2$ mm, $h_i=3$ mm, $h_g=17.5$ mm, and $h_c=30$ mm). ○, no foam; □, B–U using the airborne wave; ▽, B–U using the frame wave.

and above which the frame wave. In the middle range, say between 1000 and 2500 Hz, the average of the two curves may be used, as illustrated in Fig. 16. Compared to the B–B or U–U constructions, the B–U shell is considered to have much higher error because no single wave is dominant in the layer. However, even the lower TL curve is much higher than that of the B–B shell or double shell without the porous layer, the B–U shell will be much more effective than the B–B shell, which is a conclusion compatible with the flat panel cases.

C. U–U shell

In this case, the porous layer is installed with a 7.5 mm gap inside and a 10 mm gap outside, therefore $R_e=0.2$ m, $R_i=0.15$ m, $R_{c1}=0.159$ m, $R_{c2}=0.189$ m, $h_e=2$ mm, $h_i=3$ mm, $h_c=30$ mm, $h_{g1}=7.5$ mm, and $h_{g2}=10$ mm. In this case, the airborne wave is selected as the only wave component to represent the porous material as an equivalent fluid (This should come for each construction). Figure 17 compares the TL calculated U–U shell with that of the double walled shell without any porous layer. As it is seen in the figure, adding the porous layer in a U–U configuration substantially increases the TL over the double walled shell without the porous layer.

D. Discussions

Figure 18 compares the TLs of the foam lined cylindrical double walled shell calculated for the three configurations, B–B, B–U, and U–U. Notice that the B–U and U–U configurations show much higher TL compared to the B–B shell. Generally the TL curve of the B–U shell lies in between the two TL curves corresponding to the B–U shell in two different approximations, however closer to the higher TL one of the two B–U curves. Therefore, the B–U construction can be considered as the best for curved double shells also, as it was in the case of the flat double panels.

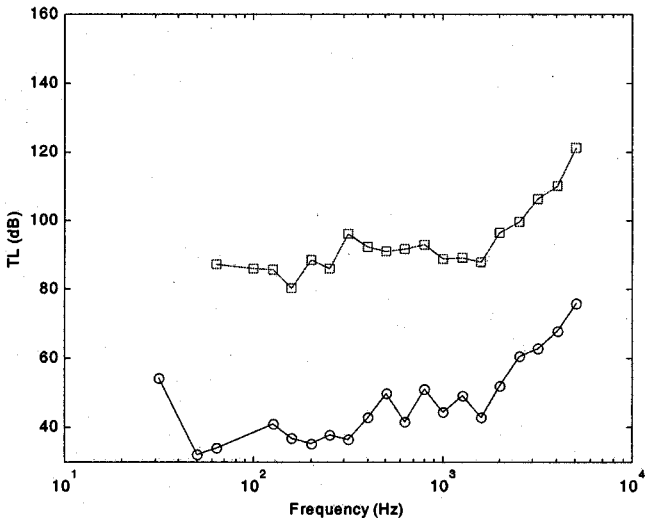


FIG. 17. A comparison of calculated TLs of cylindrical double walled shells ($R_e=0.2$ m, $R_i=0.15$ m, $R_{c1}=0.159$ m, $R_{c2}=0.189$ m, $h_e=2$ mm, $h_i=3$ mm, $h_c=30$ mm, $h_{g1}=7.5$ mm, and $h_{g2}=10$ mm). ○, no foam; □, U–U.

The simplification that uses only one dominant carrier to model the porous layer certainly sacrifices the accuracy in solutions, however, makes a generally unsolvable problems by the state of the art solvable. Obviously the intended use of the approximate method is to help qualitative or comparative design analysis or design parameter studies.

V. CONCLUSION

A simplified analysis method was proposed to solve sound transmission through structures with porous liners. The analysis method developed by Boton *et al.*[2] using the multiwave description of the porous material based on Biot's theory[1] is utilized as a preprocessor in the proposed method. Biot's theory[1] views the porous material as a homogeneous aggregate of the elastic frame and fluid pores, and therefore describes the material using three waves: two dilatational and one rotational wave. The simplified method uses a very simple concept, keeping only the strongest wave among the three to model the porous material, which is essentially modeling the porous layer as an equivalent fluid layer. For example, double panels of various core constructions are modeled as a double panel with one (B–B model), two (B–U model), or three (U–U model) layers of the fluids. The procedure to apply the method to real structures with arbitrary geometry and boundary conditions is composed of two steps. In the first step an equivalent flat double panel problem of infinite extent is solved by the full method, keeping the same cross sectional geometry to identify the strongest wave by comparing the energies associated with the waves. In the second step the problem is modeled using the actual geometry but modeling the porous layer as the equivalent fluid layer, which can be solved much more easily.

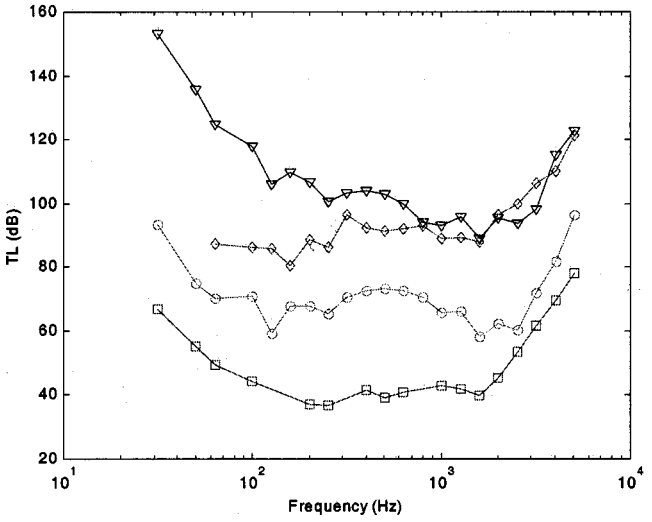


FIG. 18. Summary of calculated TLs of the three double walled shells ($R_e=0.2$ m, $R_i=0.15$ m, $h_e=2$ mm, and $h_i=3$ mm). □, B–B; ▽, B–U using the airborne wave; ○, B–U using the frame wave; ◇, U–U.

Because the proposed approximation takes simplicity of analysis at the expense of the accuracy, the accuracy (or inaccuracy) of the approximate method was checked by comparing the results with those from the full model and measurements reported in Ref. 2. The comparison shows that the simplified model provides very reasonable results compared to the full model. As application examples, the simplified method developed is applied to obtain the TLs through double walled cylindrical shells with three types of porous cores (BB, BU, and UU). The approach allows analytic solutions to the problems, which will not be possible otherwise. The calculation results show that the BU and UU shells show much higher TLs compared to the BB shell, which is in line with the flat double panel case.

Because the proposed method reduces the porous material to a fluid, most of the widely available analysis methods, i.e., FEM software with standard acoustic analysis capability, can be used without any modifications. The obvious downsides may be considered the loss of accuracy and theoretical exactness, which are not considered significant if it is considered that the full method also involves quite a few idealizations such as homogeneity and isotropy, and the property of the porous material has inherently wider variations depending on its manufacturing and installation procedures. On the other hand, the potential gain of using the proposed method is significant: sound transmission through structures with a porous material liner, which has any realistic geometry, can be analyzed.

ACKNOWLEDGMENT

The first two authors acknowledge the financial support by ArvinMeritor Industries related to this work.

[1] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 168–191 (1956).
[2] J. S. Bolton, N.-M. Shiau, and Y. J. Kang, "Sound Transmission Through Multi-Panel Structures Lined With Elastic Porous Materials," J. Sound Vib. **191**, 317–347 (1996).
[3] J. W. S. Rayleigh, *The Theory of Sound*, 2nd ed. (Dover, New York, 1945; re-issue, 1986), Vol. II, article 351.
[4] B. Brouard, D. Lafarge, and J-F Allard, "A general method of modeling sound propagation in layered media, absorption of sound by porous walls," J. Sound Vib. **183**, 129–142 (1995).
[5] K. Attenborough, "Acoustical characteristics of porous materials," Phys. Rep. **82**, 179–227 (1982).

[6] J. S. Bolton and E. R. Green, ''Normal incidence sound transmission through double-panel systems lined with relatively stiff, reticulated polyurethane foam,'' Appl. Acoust. **39**, 23–51 (1993).
[7] J. S. Bolton, ''Cepstral techniques in the measurement of acoustic reflection coefficients, with application to the determination of acoustic properties of elastic porous materials,'' Ph.D. thesis, Institute of Sound and Vibration Research, University of Southampton, 1984.
[8] J. A. Moore and R. H. Lyon, ''Resonant porous material absorbers,'' J. Acoust. Soc. Am. **72**, 1989–1999 (1982).
[9] C. Bruer and J. S. Bolton, ''Vibro-acoustic damping of extended vibration systems,'' *11th AIAA Aeroacoustics Conference*, Sunnyvale, CA, 19–21 October 1987, paper AIAA-87-2661, 1987.
[10] Y. J. Kang and J. S. Bolton, ''A finite element model for sound transmission through foam-lined double-panel structures,'' J. Acoust. Soc. Am. **99**, 2755–2765 (1996).
[11] A. Craggs, ''A finite element model for rigid porous absorbing materials,'' J. Sound Vib. **61**, 101–111 (1978).
[12] V. Easwaran, W. Lauriks, and L. Kelders, ''Finite element analysis of axial wave propagation in poro-elastic media,'' *Proceedings of Inter-Noise 95* (Noise Control Foundation, New York, 1995), Vol. 1, pp. 445–448.
[13] K. Attenborough, ''The influence of microstructure on propagation in porous fibrous absorbents,'' J. Sound Vib. **16**, 419–442 (1971).
[14] C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, Amsterdam, 1949).
[15] J. F. Allard, A. Aknine, and C. Depollier, ''Acoustical properties of partially reticulated foams with high and medium flow resistance,'' J. Acoust. Soc. Am. **79**, 1734–1140 (1986).
[16] J. F. Allard, C. Depollier, and W. Lauriks, ''Measurement and prediction of surface impedance at oblique incidence of a plastic foam of high flow resistivity,'' J. Sound Vib. **132**, 51–60 (1989).
[17] L. L. Beranek, ''Acoustical properties of homogeneous, isotropic rigid tiles and flexible blankets,'' J. Acoust. Soc. Am. **19**, 556–568 (1947).
[18] L. L. Beranek and G. A. Work, ''Sound transmission through multiple structures containing flexible blankets,'' J. Acoust. Soc. Am. **21**, 419–428 (1949).
[19] K. A. Mulholland, A. J. Price, and H. D. Parbrook, ''Transmission loss of multiple panels in a random incidence field,'' J. Acoust. Soc. Am. **43**, 1432–1435 (1968).
[20] J. S. Bolton and N.-M. Shiau, ''Oblique incidence sound transmission through multi-panel structures lined with elastic porous material,'' in Ref. 9, paper AIAA-86-2660, 1987.
[21] J. S. Bolton and N.-M. Shiau, ''Random incidence transmission loss of lined, finite double panel systems,'' *12th AIAA Aeroacoustics Conference*, San Antonio, TX, 10–12 October 1989, paper AIAA-89-1048.
[22] J. P. Coyette and H. Wynendaele, ''A finite element model for predicting the acoustic transmission characteristics of layered structures,'' in Ref. 12, 1995, pp. 1279–1282.
[23] T. E. Vigran, L. Kelders, W. Lauriks, M. Dhainaut, and T. F. Johansen, ''Force response of a sandwich plate with flexible core described by Biot model,'' Acta Acoustica **83**, 1024–1031 (1997).
[24] L. Cremer, M. Heckl, and E. E. Ungar, *Structure-Borne Sound* (Springer-Verlag, Berlin, 1973).
[25] A. D. Pierce, *Acoustics* (McGraw-Hill, New York, 1981).
[26] K. A. Mulholland, H. D. Parbrook, and A. Cummings, ''The transmission loss of double panels,'' J. Sound Vib. **6**, 324–334 (1967).
[27] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982).
[28] W. Soedel, *Vibrations of Shells and Plates* (Marcel Dekker, New York, 1993).
[29] M. Morse and K. Uno Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[30] C.-Y. Wang and R. Vaicaitis, ''Active control of vibrations and noise of double wall cylindrical shells,'' J. Sound Vib. **216**, 865–888 (1998).
[31] L. R. Koval, ''Sound transmission into a laminated composite cylindrical shell,'' J. Sound Vib. **71**, 523–530 (1980).
[32] R. Vaicaitis and D. A. Bofilios, ''Noise transmission of double wall composite shells,'' AIAA Aerospace Science Meeting and Exhibit, paper No. 85-0786-CP, 1986.
[33] D. A. Bofilios and R. Vaicaitis, ''Response of double-wall composite shells to random point loads,'' J. Aircr. **24**, 268–273 (1987).
[34] A. Blaise, C. Lesuer, M. Gotteland, and M. Barbe, ''On sound transmission into an orthotropic infinite shell: Comparison with Koval's results and understanding of phenomena,'' J. Sound Vib. **150**, 233–243 (1991).
[35] Y. Y. Tang, J. H. Robinson, and R. J. Silcox, ''Sound transmission through a cylindrical sandwich shell with honeycomb core,'' *34th AIAA Aerospace Science Meeting and Exhibit*, AIAA-96-0877, 1996, pp. 1–10.
[36] Y. Y. Tang, R. J. Silcox, J. H. Robinson, ''Sound transmission through two concentric cylindrical sandwich shells,'' *Proceedings of 14th International Modal Analysis Conference*, 1996, pp. 1488–1492.
[37] C. L. Dym and M. A. Lang, ''Transmission of sound through sandwich panels,'' J. Acoust. Soc. Am. **56**, 1523–1532 (1974).
[38] F. W. Grosveld and K. P. Shepherd, ''Active sound attenuation across a double wall structure,'' J. Aircr. **31**, 223–227 (1994).
[39] J. H. Lee and J. Kim. ''Analysis and measurement of sound transmission through a double-walled cylindrical shell,'' J. Sound Vibr. (in press).
[40] J. H. Lee and J. Kim, ''Sound transmission through double-walled cylindrical shells sandwiching porous material,'' *Proceedings of the 2001 SAE Noise and Vibration Conference*, 2001-01-1518, Traverse City, MI, May 2001.

Coupled guided acoustic modes in water-filled thin-walled tubes

G. Maze
LAUE UMR CNRS 6068, Université du Havre, Place Robert Schuman, 76610 Le Havre, France

J. D. N. Cheeke
Physics Department, Concordia University, 1455 de Maissonneuve Boulevard West, Montréal, Québec H3G 1M8, Canada

X. Li
Nortel Networks, Nepean, Ontario Canada

Z. Wang
Physics Department, Concordia University, 1455 de Maissonneuve Boulevard West, Montréal, Québec H3G 1M8, Canada

(Received 13 August 2000; revised 19 July 2001; accepted 27 July 2001)

Recent results [J. D. N. Cheeke *et al.*, J. Acoust. Soc. Am. **104**, 3678 (1998)] reported a 20% decrease in group velocity for flexural mode circumferential waves propagating around water-filled thin-walled stainless-steel tubes. In the present work, the full theory of such modes is developed to explain the structure modes which originates from the coupling between the water filler and stainless tube. Calculated values of the group velocity for the first two coupled modes are in excellent agreement with experiment. The results are of interest for the physics of acoustic waves in fluid-loaded structures and have potential application for liquid level detection. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1406161]

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

I. INTRODUCTION

Recently there has been considerable interest in the understanding of acoustic modes in thin plates and thin-walled cylindrical shells either liquid loaded on one side or immersed in fluids.[1–16] For one-sided loading a Scholte/Stoneley interface wave is formed, and it has been shown that there is a repulsion phenomenon between the dispersion curve of this mode[6–8] and that of the A_0 mode of the unloaded structure. This occurs near a critical frequency where the phase velocity of the A_0 mode approaches that of the liquid. In addition, there is a change in mode character between A_0 and A modes above and below the critical frequency for both plates[5] and tubes.[7,8] The A modes are detected only for a thin cylindrical shell ($b/a>0.7$),[6] it is also called A_0^- wave. Other authors[9–11] showed that for a tube immersed in a fluid with a ''filler'' liquid inside, the latter contains new modes SGWm which also display repulsion and veering effects with the A mode for a thin shell[9,10] or with the A_0 mode for a thick shell. The present work shows new theoretical results for the case of a fluid-filled thin-walled cylindrical shell in vacuum. Following development of the theory in Sec. II, this is compared to recent experimental results[14] in Sec. III; excellent agreement is obtained.

II. THEORY

The particular case under consideration is that of a water-filled, thin-walled stainless-steel cylindrical shell with $b/a=0.95$ ($a=$outer, $b=$inner radius, wall thickness $d=a-b=0.25$ mm). The material parameters are given in the Appendix. For simplicity the shell is considered to be in vacuum.

To describe the vibrating state in a plane perpendicular to the cylindrical shell axis (Fig. 1), the scalar potential and the axial component of the vector potential in the shell, respectively, in relation to the longitudinal and shear vibration, can be written for each mode n (n: integer)

$$\varphi_{2n}=i^n\varepsilon_n[a_nJ_n(k_lr)+b_nN_n(k_lr)]\cos(n\theta)\exp(-i\omega t),$$

$$\psi_{2n}^{(z)}=i^n\varepsilon_n[c_nJ_n(k_tr)+d_nN_n(k_tr)]\sin(n\theta)\exp(-i\omega t).$$

The scalar potential in the water column is

$$\varphi_{3n}=i^n\varepsilon_n[e_nJ_n(k_wr)]\cos(n\theta)\exp(-i\omega t),$$

with

$$k_l=\omega/C_l,\quad k_t=\omega/C_t,\quad k_w=\omega/C_w,$$

$\boldsymbol{\varepsilon}_\mathbf{n}$ is the Neumann coefficient, $\boldsymbol{\varepsilon}_\mathbf{n}=1$ if $n=0$ and $\boldsymbol{\varepsilon}_\mathbf{n}=2$ if $n\neq 0$. $\omega=2\pi f$ is the acoustic wave frequency, r is radial, and θ angular coordinates.

To solve this problem for five unknown quantities (a_n,b_n,c_n,d_n,e_n) the boundary conditions are (see the Appendix):

(1) for $r=a$: (i) The radial stress is equal to 0; and (ii) the tangential stress is equal to 0;
(2) for $r=b$: (i) The radial components of the displacement in the shell and in water are identical; (ii) the radial stress in the shell and in water are identical; and (iii) the tangential stress is equal to 0.

The problem can then be written

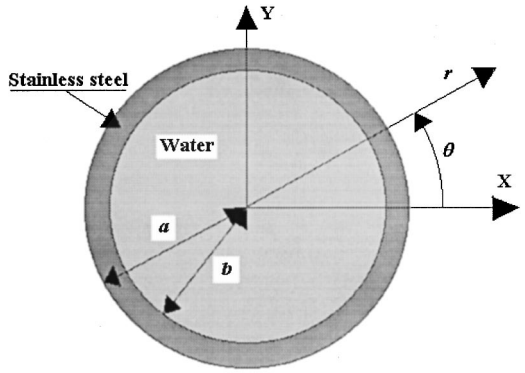


FIG. 1. Cross-sectional view of the shell configuration.

$$[\mathrm{M}]\begin{bmatrix} a_n \\ b_n \\ c_n \\ d_n \\ e_n \end{bmatrix} = 0,$$

where [M] is a 5×5 matrix given in the Appendix. This problem has a solution if the determinant $|\mathrm{M}|$ is equal to 0. The roots x_n^* are eigenmodes of this problem; they are also called resonances (Table I, columns 6 and 7). These values are plotted on the Regge curves $n=f(x)$ of Fig. 2 (-■-,-□-). Only two trajectories are calculated to compare with experiment (for each mode n two values x_n^* are obtained); in reality there exist an infinity of values. Each trajectory is related to a circumferential wave, where n is the number of wavelengths around the circumference of the object.

With these results, the dispersion curves of the phase velocity are calculated applying the following relation for each Regge trajectory:

$$C_L^{ph}=C_w\frac{x_{Ln}^*}{n},$$

where x_{Ln}^* is the dimensionless frequency of an eigenmode **n** of the trajectory L ($L=1$ or 2). The results are displayed in Fig. 3. The dispersion curves of the group velocity are calculated using the relation

$$V_L^{gr}=C_w(x_{L,i+1}^*-x_{L,i}^*),$$

where $i+1$ and i indicate two successful eigenmodes. The results are displayed in Fig. 4.

To explain the shape of these curves, the eigenmodes related to the flexural Lamb wave (A_0) of the empty cylindrical shell in vacuum have been calculated (Table I, column 1) and are shown in Figs. 2, 3, and 4 (-●-). The first two trajectories of the eigenmodes of a column of water situated in a rigid circular surrounding wall are shown in the same figures (-×-, -+-) (see Table I, columns 3 and 4). The mode order (first, second,...) is defined with respect to the value of the corresponding resonance frequency for a given mode n.

From Fig. 2 and Table I the following facts are observed:

(i) The curve of the eigenmodes of the empty tube in vacuum related to the flexural A_0 wave crosses the curves of the eigenmodes of the water column.

(ii) The eigenmodes of water-filled tube in vacuum are real, the imaginary part is null. The energy re-emitted in water filling the cavity is used to re-excite the modes. The energy is not scattered outside the water-filled tube.

(iii) At low frequencies ($x<20$) (Table I) the first curve of the eigenmodes of the water-filled tube (-■-) is close to the curve of the flexural A_0 wave (-●-), which means that the circumferential wave around the water-filled tube is identical to the flexural wave. In this case, the eigenmodes of the structure depend principally on the shell parameters. At high frequencies ($x>40$), the curve of the first eigenmodes of the water-filled tube is close to the curve of the first eigenmodes of the water column (-×-); the eigenmodes of the structure depend principally on the water cylinder parameters. In intermediate frequencies the eigenmodes depend on the entire structure.

(iv) At low frequencies ($x<30$) (Table I) the second curve of the eigenmodes of the water-filled tube (-□-) is close to the first curve of the eigenmodes of the water column (-×-); they depend on the water parameters. At the middle frequencies ($35<x<45$) it is close to the curve of the A_0 wave (-●-); the influence of shell is strong, and at high frequencies ($x>50$) it is close to the second curve of the water column (-+-); the parameters of the water cylinder play a leading part. This shows that the character of the circumferential wave changes over the frequency range considered.

In Table I, the theoretical complex modes of A wave experimentally observable for an empty cylindrical shell immersed in water are noted (column 5) ($12\leqslant n\leqslant 23$). In this case, the imaginary part ($\Gamma/2$) is related to the energy scattered in the water surrounding the tube and this energy cannot be used to re-excite the modes. In Fig. 5, the real part (-○-) can be compared to the dimensionless frequencies of the first serie of eigenmodes of the water-filled tube (-■-). (Table I, column 5). These modes are close but the gap between two n modes increases with n when the influence of the water column increases.

The dispersion curve of the phase velocity (Fig. 3) confirms the above remarks. The velocity at low frequencies (-■-) is close to that of the flexural wave A_0 in the cylindrical shell, and at high frequencies is close to the velocity of the lowest mode in the water column. When the frequency increases this velocity tends to the velocity of sound in water ($C_w=1470$ m/s). The velocity curve (-□-) at low frequencies is parallel to that of the velocity of the first wave in the water column, in middle frequencies it is close to the velocity curve of the flexural wave, and at high frequencies it is close to the velocity of the second wave in the water column.

III. RESULTS AND DISCUSSION

The experiments on thin-walled stainless steel (SS) tubes were carried out using interdigital transducer (IDTs) overlaying PZT films deposited by sol-gel as described previously.[13] The tubes had an outer diameter of 9.8 mm and wall thickness of 0.254 mm. PZT films were about 5 μm thick, and in the results to be shown here a chirp IDT was used so that circumferential flexural modes, particularly A_0, could be studied over the range 0.8–6 MHz by simply

TABLE I. Frequencies of modes of A_0 wave of an empty shell in vacuum (2); frequencies of first modes of water column (3); frequencies of second modes of water column (4); complex modes of A wave of an empty shell immersed in water (5); frequencies of first modes of water-filled shell in vacuum (6); frequencies of second modes of water-filled shell in vacuum (7).

1 Mode n	2 Modes of A_0 wave of a shell in vacuum	3 First modes of water column	4 Second modes of water column	5 Modes of A wave of an empty shell immersed in water	6 First modes of water-filled shell in vacuum	7 Second modes of water-filled shell in vacuum
1		1.95	5.62		⋯	2.45
2	0.15	3.22	7.06		⋯	3.95
3	0.41	4.43	8.44		⋯	5.18
4	0.79	5.60	9.78		0.63	6.34
5	1.26	6.76	11.07		1.04	7.48
6	1.84	7.91	12.36		1.56	8.61
7	2.52	9.03	13.62		2.18	9.73
8	3.29	10.16	14.86		2.89	10.85
9	4.16	11.28	16.09		3.69	11.96
10	5.11	12.40	17.32		4.57	13.08
11	6.14	13.51	18.54		5.54	14.19
12	7.26	14.61	19.74	6.49–i 0.000 07	6.58	15.31
13	8.45	15.72	20.94	7.49–i 0.000 19	7.70	16.42
14	9.71	16.82	22.13	8.75–i 0.000 51	8.89	17.54
15	11.03	17.93	23.32	9.96–i 0.001 4	10.13	18.66
16	12.42	19.02	24.49	11.23–i 0.003 6	11.44	19.79
17	13.87	20.12	25.67	12.54–i 0.007 5	12.80	20.92
18	15.38	21.21	26.84	13.88–i 0.021 8	14.21	22.05
19	16.94	22.31	28.01	15.24–i 0.050 0	15.65	23.20
20	18.54	23.39	29.18	16.62–i 0.105 9	17.14	24.35
21	20.19	24.48	30.34	17.99–i 0.214 1	18.65	25.51
22	21.88	25.57	31.49	19.36–i 0.408 0	20.19	26.69
23	23.61	26.66	32.65	20.67–i 0.746 5	21.75	27.89
24	25.38	27.75	33.80		23.31	29.12
25	27.18	28.83	34.96		24.87	30.37
26	29.01	29.92	36.11		26.41	31.66
27	30.87	31.00	37.25		27.93	33.00
28	32.76	32.08	38.39		29.41	34.38
29	34.67	33.17	39.54		30.85	35.82
30	36.61	34.25	40.67		32.25	37.31
31	38.56	35.34	41.81		33.39	38.83
32	40.54	36.41	42.95		34.89	40.38
33	42.53	37.49	44.08		36.15	41.93
34	44.54	38.58	45.22		37.38	43.46
35	46.57	39.65	46.35		38.58	44.93
36	48.61	40.74	47.48		39.76	46.34
37	50.66	41.81	48.61		40.92	47.69
38	52.73	42.88	49.74		42.07	48.98
39	54.81	43.97	50.86		43.21	50.22
40	56.89	45.04	51.98		44.33	51.44
41	58.99	46.12	53.12		45.45	52.63
42	61.09	47.20	54.24		46.57	53.81
43	63.21	48.27	55.36		47.67	54.97
44	65.33	49.35	56.48		48.78	56.13
45	67.45	50.42	57.60		49.88	57.28
46	69.58	51.49	58.73		50.97	58.42
47	71.72	52.57	59.84		52.07	59.56
48	73.86	53.64	60.96		53.16	60.69
49	76.01	54.72	62.07		54.25	61.82
50	78.16	55.79	63.19		55.33	62.95
51	80.32	56.86	64.31		56.42	
52	82.48	57.94	65.42		57.50	
53	84.64	59.01	66.54		58.58	
54	86.80	60.08	67.64		59.66	
55	88.97	61.16	68.76		60.74	
56	91.14	62.23	69.87		61.82	
57	93.31	63.29	70.98		62.90	
58	95.48	64.37	72.09		63.97	
59	97.66	65.44			65.05	
60	99.83	66.52			66.12	

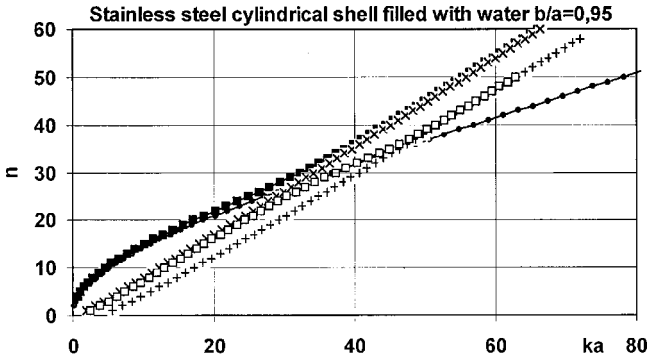


FIG. 2. Regge plot of mode number **n** as a function of dimensionless frequency x, -■- trajectory of the first eigenmodes of the water-filled tube, -□- trajectory of the second eigenmodes of the water-filled tube, -●- trajectory of the eigenmodes of flexural A_0 wave, -×- trajectory of the first eigenmodes of the water column, -+- trajectory of the second eigenmodes of the water column.

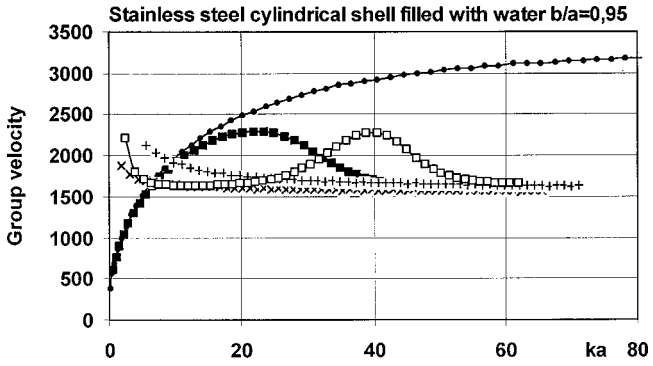


FIG. 4. Group velocity as a function of dimensionless frequency $x=ka$ -■- for the first eigenmodes of the water-filled tube, -□- for the second eigenmodes of the water-filled tube, -●- for the eigenmodes of flexural A_0 wave, -×- for the first eigenmodes of the water column, -+- for the second eigenmodes of the water column.

sweeping the frequency during a single experiment.[14] This corresponds to the range $fd=0.20$ to 1.5 MHz mm, which covers the critical frequency area mentioned earlier. Pulse-echo measurements were made using a Ritec 10000 system to produce tone bursts for which the chip transducer acted as generator and receiver. Thus, group velocities can be obtained directly from the time internal between received echoes knowing the radius of the shell. The phase velocity curve was then calculated by integrating the group velocity dispersion curve as described elsewhere.[14] V_g and V_p curves were obtained over the full frequency range for the empty tube and they are in reasonable agreement with theory.

The results for the SS tube are shown in Fig. 3 of Ref. 14. For the water-filled tube it was only possible to observe echoes over the range 0.8–2.4 MHz, with a small gap in the middle of this interval as seen from the figure. In order to distinguish between leaky waves and Stoneley waves as the source of the echoes, various obstacles such as concentric tubes, planar reflectors, etc., were placed in the water-filled tube. No change was seen in the observed echoes. As discussed in detail previously[15] this implies that the echoes are not due to leaky waves.

The central question is to identify the experimentally observed modes. Clearly, in comparing with the theoretical group velocities, we are led to associate the two halves of the plateau with the theoretical predicted fairly broad maxima of V_g for the two modes of the water-filled tube. Several questions remain, however, and only partial answers can be given here. Experimentally, the echoes disappear in the gap region of the plateau. From the theoretical results it is not clear that this should be so, and further work is needed on this point. To explain this particular result, it would be necessary to calculate the components of the vibrating vector on the outer surface of tube; the experimental setup cannot detect all polarizations.

A more subtle issue is why we are in fact able to observe an interface mode in this configuration where the signal is detected in the outer tube wall. In the case of the A wave circumnavigating around an empty cylindrical shell immersed in water with a radius ratio b/a superior to 0.7, the resonance modes are experimentally detected in the frequency interval where this wave begins to deviate from the A_0 wave of empty tube in vacuum. This critical zone is related to the crossing of the A_0 modes trajectory of an empty cylindrical shell in vacuum and the first Franz mode trajectory of rigid cylinder,[6] the coupling between this A wave and the water is not too small and not too large; the imaginary part of the resonance modes calculated is intermediate. The

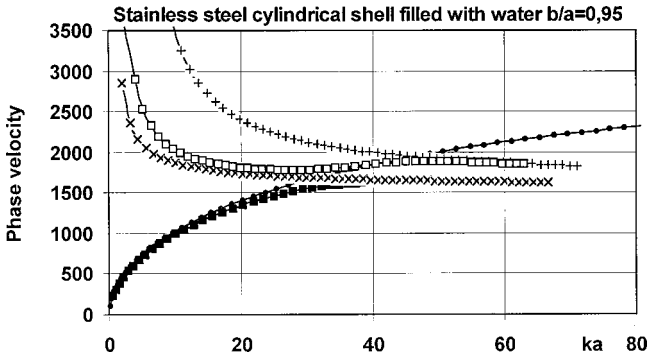


FIG. 3. Phase velocity as a function of dimensionless frequency $x=ka$ -■- for the first eigenmodes of the water-filled tube, -□- for the second eigenmodes of the water-filled tube, -●- for the eigenmodes of flexural A_0 wave, -×- for the first eigenmodes of the water column, -+- for the second eigenmodes of the water column.

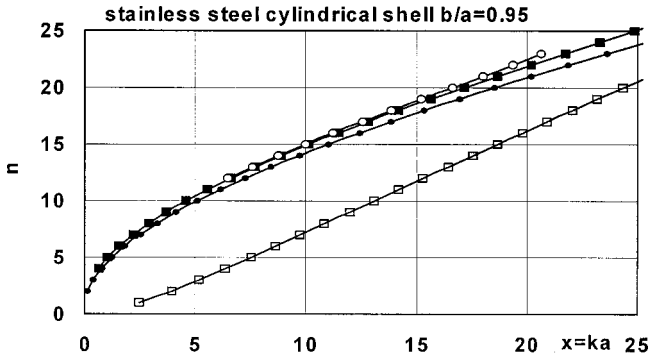


FIG. 5. Comparison between the real parts of the A mode n (-○-) and (i) the dimensionless frequencies of A_0 modes of empty tube in vacuum (-●-); (ii) the dimensionless frequencies of the first modes of water-filled tube (-■-); (iii) the dimensionless frequencies of the second modes of water-filled tube (-□-).

normal component of the vibrating vector is not preponderant. If the polarization of the vibrating vector is tangential, the coupling is null, and if the polarization is only normal, the coupling is too strong and the interface wave cannot propagate; the energy is instantaneously re-emitted in water. It is known that for a dimensionless frequency above the critical region the *A* mode is mainly water-like in character; the asymptotic limit is the Franz trajectory, while well below it the wave is mainly in the shell, the asymptotic trajectory is the A_0 trajectory of the empty shell in vacuum.

In the case of this study, the structure waves are detected when the group velocity is maximum. For the first structure wave, the maximum corresponds to the frequency zone where the mode trajectory begins to deviate from the A_0 mode trajectory, the wave is always essentially a shell wave; above the critical zone the structure wave has water parameters. For the second structure wave the maximum corresponds to mode trajectory close to the A_0 trajectory; this wave is without doubt a shell wave; below and above the critical zone this second structure wave has water parameters. Without further knowledge it is reasonable to assume that in the repulsion region the wave is uniformly shared by both fluid and tube, which would reasonably explain our ability to detect it in the tube.

In conclusion, the present work gives a convincing theoretical and experimental demonstration of the existence of the structure waves inside a water-filled, thin-walled stainless-steel tube in vacuum. These waves are experimentally detected when a strong interaction establishes in the structure between a shell wave and a water column wave. As for the *A* wave in a shell immersed in water, these waves are detected when the group velocity has a maximum value, the derivative is stationary. The first structure wave is probably similar to the *A* wave detected with an empty cylindrical shell immersed in water. In a complex structure it is not possible to talk of a shell wave or water wave, but it is necessary to talk of a structure wave. Further work will be needed to elucidate the detailed nature of this wave.

ACKNOWLEDGMENT

The experimental part of this work was supported by the Natural Sciences and Engineering Research Council of Canada.

APPENDIX

1. Variable definitions

a: outer radius of the stainless steel cylindrical shell;
b: inner radius of the stainless steel cylindrical shell;
$C_w = 1470 m/s$ sound velocity in water;
$C_l = 5790 m/s$: velocity of the longitudinal wave in stainless steel;
$C_t = 3100 m/s$: velocity of the shear wave in stainless steel;
$\rho_2 = 7900 \text{kg/m}^3$: density of stainless steel;
$\rho_3 = 1000 kg/m^3$: density of water.
x is the dimensionless frequency
$x = k_w a = 2\pi a f / C_w$, f: ultrasonic frequency.
$x_l = x\, C_w / C_l$, $x_t = x\, C_w / C_t$,
$y_l = x_l b/a$, $y_t = x_t b/a$, $y_3 = xb/a$.

2. Equations of continuity

(i) For $r = a$
The radial stress is equal to 0:

$$\lambda_2 \Delta \phi_{2n} + 2\mu_2 \left[\frac{\partial^2 \phi_{2n}}{\partial r^2} + \frac{1}{r} \frac{\partial^2 \psi_{2n}^{(z)}}{\partial r \partial \theta} - \frac{1}{r^2} \frac{\partial \psi_{2n}^{(z)}}{\partial \theta} \right] = 0.$$

The tangential stress in a plan perpendicular to axial axis is equal to 0:

$$2\left[\frac{1}{r} \frac{\partial^2 \phi_{2n}}{\partial r \partial \theta} - \frac{1}{r^2} \frac{\partial \phi_{2n}}{\partial \theta} \right] + \left[\frac{1}{r^2} \frac{\partial^2 \psi_{2n}^{(z)}}{\partial \theta^2} - \frac{\partial^2 \psi_{2n}^{(z)}}{\partial r^2} + \frac{1}{r} \frac{\partial \psi_{2n}^{(z)}}{\partial r} \right] = 0.$$

(ii) For $r = b$
The radial components of displacement in water and in shell are equal:

$$\frac{\partial \phi_{2n}}{\partial r} + \frac{1}{r} \frac{\partial \psi_{2n}^{(z)}}{\partial \theta} = \frac{\partial \phi_{3n}}{\partial r}$$

The radial stress in the shell is equal to the pressure in water:

$$\lambda_2 \Delta \varphi_{2n} + 2\mu_2 \left[\frac{\partial^2 \varphi_{2n}}{\partial r^2} + \frac{1}{r} \frac{\partial^2 \psi_{2n}^{(z)}}{\partial r \partial \theta} - \frac{1}{r^2} \frac{\partial \psi_{2n}^{(z)}}{\partial \theta} \right] = \lambda_3 \Delta \phi_{3n}.$$

The tangential stress in a plan perpendicular to axial axis is equal to 0:

$$2\left[\frac{1}{r} \frac{\partial^2 \phi_{2n}}{\partial r \partial \theta} - \frac{1}{r^2} \frac{\partial \phi_{2n}}{\partial \theta} \right] + \left[\frac{1}{r^2} \frac{\partial^2 \psi_{2n}^{(z)}}{\partial \theta^2} - \frac{\partial^2 \psi_{2n}^{(z)}}{\partial r^2} + \frac{1}{r} \frac{\partial \psi_{2n}^{(z)}}{\partial r} \right] = 0.$$

3. Determinant

$$M_n = |M_{ij}| \qquad (i,j = 1-5),$$

with n vibration mode (integer).
Coefficients of the determinant:

$$M_{11}=(-x_t^2+2n^2)J_n(x_l)-2x_lJ_n'(x_l)$$
$$M_{12}=(-x_t^2+2n^2)N_n(x_l)-2x_lN_n'(x_l)$$
$$M_{13}=2n[x_tJ_n'(x_t)-J_n(x_t)]$$
$$M_{14}=2n[x_tN_n'(x_t)-N_n(x_t)]$$
$$M_{15}=0$$
$$M_{21}=2n[-x_lJ_n'(x_l)+J_n(x_l)]$$
$$M_{22}=2n[-x_lN_n'(x_l)+N_n(x_l)]$$
$$M_{23}=(x_t^2-2n^2)J_n(x_t)+2x_tJ_n'(x_t)$$
$$M_{24}=(x_t^2-2n^2)N_n(x_t)+2x_tN_n'(x_t)$$
$$M_{25}=0$$
$$M_{31}=(y_t^2-2n^2)J_n(y_t)+2y_tJ_n'(y_t)$$
$$M_{32}=(y_t^2-2n^2)N_n(y_t)+2y_tN_n'(y_t)$$
$$M_{33}=2n[y_tJ_n'(y_t)-J_n(y_t)]$$
$$M_{34}=2n[y_tN_n'(y_t)-N_n(y_t)]$$
$$M_{35}=\frac{\rho_3}{\rho_2}y_t^2J_n(y_3)$$
$$M_{41}=y_lJ_n'(y_l)$$
$$M_{42}=y_lN_n'(y_l)$$
$$M_{43}=nJ_n(y_t)$$
$$M_{44}=nN_n(y_t)$$
$$M_{45}=-y_3J_n'(y_3)$$
$$M_{51}=2n[-y_lJ_n'(y_l)-J_n(y_l)]$$
$$M_{52}=2n[-y_lN_n'(y_l)-N_n(y_l)]$$
$$M_{53}=(y_t^2-2n^2)J_n(y_t)+2y_tJ_n'(y_t)$$
$$M_{54}=(y_t^2-2n^2)N_n(y_t)+2y_tN_n'(y_t)$$
$$M_{55}=0$$

[1] H. Überall, "Surface waves in acoustics," Vol. X in *Physical Acoustics*, edited by W. D. Mason and R. N. Thurson (Academic Press, New York, 1973).

[2] H. Überall, B. Hosten, M. Deschamps, and A. Gérard, "Repulsion of phase-velocity dispersion curves and the nature of plate vibrations," J. Acoust. Soc. Am. **96**, 908–917 (1994).

[3] J. Dickey, G. Maidanik, and H. Überall, "The splitting of dispersion curves for fluid-loaded plates," J. Acoust. Soc. Am. **98**, 2365–2367 (1995).

[4] J. P. Sessarego, J. Sageloli, C. Gazanhes, and H. Überall, "Two Scholte–Stoneley waves on doubly fluid-loaded plates and shells," J. Acoust. Soc. Am. **101**, 135–142 (1997).

[5] X. L. Bao, H. Franklin, P. K. Raju, and H. Überall, "The splitting of dispersion curves for plates fluid-loaded on both sides," J. Acoust. Soc. Am. **102**, 1246–1248 (1997).

[6] G. Maze, F. Léon, J. Ripoche, A. Klauson, J. Metsaveer, and H. Überall, "Nature de l'onde d'interface de Scholte sur une coque cylindrique," Acust. Int. J. Acoust. **81**, 201–213 (1995).

[7] A. C. Ahyi, P. Pernod, O. Gatti, V. Latard, A. Merlen, and H. Überall, "Experimental demonstration of the pseudo-Rayleigh (A_0) wave," J. Acoust. Soc. Am. **104**, 2727–2732 (1998).

[8] G. Maze, F. Léon, J. Ripoche, and H. Überall, "Repulsion phenomena in the phase-velocity dispersion curves of circumferential waves on elastic cylindrical shells," J. Acoust. Soc. Am. **105**, 1695–1701 (1999).

[9] N. D. Veksler, J. M. Conoir, and J. L. Izbicki, "Fluid-filled cylindrical shell: Influence of the fluid-structure coupling on the peripheral waves," Fourth International Congress on Sound and Vibration, St. Petersburg, Russia, 24–27 June 1996.

[10] X. L. Bao, P. K. Raju, and H. Überall, "Circumferential waves on an immersed, fluid-filled elastic cylindrical shell," J. Acoust. Soc. Am. **105**, 2704–2709 (1999).

[11] N. D. Veksler, J.-L. Izbicki, and J.-M. Conoir, "Flexural waves in the acoustic wave scattering by a liquid-filled shell," Acoust. Phys. **45**(3), 279–288 (1999).

[12] M. F. M. Osborne and S. D. Hart, "Transmission, reflection and guiding of an experimental pulse by a steel plate in water," J. Acoust. Soc. Am. **17**, 1–18 (1945).

[13] X. Li, J. D. N. Cheeke, Z. Wang, C. K. Jen, M. Viens, G. Yi, and M. Sayer, "Ultrasonic thin-walled tube wave devices for sensor applications," Appl. Phys. Lett. **67**(1), 37–39 (1995).

[14] J. D. N. Cheeke, X. Li, and Z. Wang, "Observation of flexural Lamb waves (A_0 mode) on water-filled cylindrical shells," J. Acoust. Soc. Am. **104**, 3678–3680 (1998).

[15] J. D. N. Cheeke, X. Li, and Z. Wang, "Characteristics of circumferential waves in thin-walled tube acoustic devices," 1995 IEEE Ultrasonics Symposium, Seattle, WA, p. 441.

[16] J. D. N. Cheeke, K. Shannon, and Z. Wang, "Loading effects on A_0 Lamb-like waves in full and partially filled thin-walled tubes," Sens. Actuators B **59**, 180–183 (1999).

Theoretical and experimental study of the influence of the particle size distribution on acoustic wave properties of strongly inhomogeneous media

François Vander Meulen,[a] Guy Feuillard, Olivier Bou Matar, Franck Levassort, and Marc Lethiecq
LUSSI-EIVL, Rue de la Chocolaterie, 41000 Blois, France

(Received 6 November 2000; revised 24 July 2001; accepted 27 July 2001)

The ultrasonic method is particularly suitable to characterize diffusive media, as acoustic properties (velocity and attenuation) are related to the properties and concentrations of the homogeneous phase and scatterers. Thus, ultrasonic characterization can be useful in the study of sedimentation or flocculation processes, in concentration measurements, and granulometry evaluation. Many models have been developed for media where particles are very small compared to the incident wavelength. When the diameter of the particles is close to the wavelength, multiple-scattering theories have to be used to describe the propagation of waves. In this paper, the case where the ratio of wavelength to scatterer size is around unity is studied. First, the particle size distribution is taken into account in two types of multiple-scattering theories based on the effective field approximation or on the quasicrystalline approximation and theoretical results are produced. The T-matrix formalism has been used to calculate the amplitude of the wave scattered by a single sphere. The calculation of the complex wave number in the effective medium has been implemented, using in particular the Percus–Yevick equation as a spatial pair-correlation function between scatterers, and a normal particle-size distribution. The influence of these parameters is discussed. Finally, attenuation and phase velocity measurements are performed in moving suspensions of acrylic spheres in ethylene glycol, at various concentrations and for different particle-size distributions. A good agreement between the theoretical results and the measurements is found for both velocity and attenuation. These results show that the size distribution is a critical parameter to understand velocity and attenuation behavior as function of frequency and volume fraction. [DOI: 10.1121/1.1404435]

PACS numbers: 43.20.Hq [ANN]

I. INTRODUCTION

Since the pioneering work of Lord Rayleigh[1] on acoustic diffusion, the ultrasonic characterization of inhomogeneous media has been extensively studied in literature.[2,3] These studies are motivated by the variety of encountered media and by the wide range of applications. A few examples are: the study of sedimentation or flocculation processes, concentration measurements, or particle-size evaluation. The motivations of these scientific works also holds in the theory of interaction of an acoustic wave with a population of scatterers. Three characteristics are important to understand the acoustic wave behavior: the concentration of scatterers, the ratio of the scatterer diameter to the wavelengths of relevant wave modes, and the scattering amplitude function of the scatterer. The latter includes the acoustic contrast between the homogeneous phase and the scatterer. According to their values, the diffusive regime can be single or multiple. Whatever the scattering cross section and particle diameter are, if the concentration of scatterers is low, single diffusion occurs. This means that each scatterer is only exited by the incident wave. If the concentration increases, the field scattered from a particle will subsequently interact with more particles and, like for a billiard ball, the acoustic path can be much longer than the direct ballistic path. In this case, multiple-scattering theories have to be considered. The fact that the scattered waves randomly travel in the inhomogeneous media has a direct consequence on these theories. They will predict the scattering wave on an average of many configurations and not on a single one. In other words, even if the acoustic wave has a random path, the average acoustic wave on many configurations is assumed to contain coherent information on multiple scattering. Thus, the effective wave number, from which phase velocity and attenuation are deduced, is also calculated for an average configuration. In practice, the signal received on a transducer must be an average of many signals corresponding to different geometrical configurations.

As it can be seen, the theoretical frame of multiple-scattering theories is well defined. However, the size distribution of scatterers is another characteristic of particular importance that must be taken into account. Indeed, the size of a scatterer defines the amplitude of the scattered wave, so the size acts as a weighting factor. The purpose of this paper is to study the influence of this distribution on multiple-scattering results. In Sec. II, the main results of some multiple-scattering theories based on the effective field approximation[4,5] or on the quasicrystalline approximation[6–9] in which particle-size distributions have been included, are produced. The T-matrix formalism has been used to calculate

[a]Electronic mail: vandermeulen@univ-tours.fr

the amplitude of the wave scattered by a single sphere. The complex wave number in the effective medium is calculated using the Percus–Yevick equation as a spatial pair-correlation function between scatterers. In the case of a normal particle-size distribution, the influence of the statistic parameters is discussed. Then, the experimental setup and the methods used to measure phase velocity and attenuation are reported. Finally, theoretical and experimental results on acrylic beads in ethylene glycol solution are given and discussed.

II. MULTIPLE SCATTERING THEORY

The starting point to describe the propagation of acoustic waves in two phase media is the motion equation of the velocity potential in the homogeneous fluid. According to the adiabatic, Euler, and mass conservation equations, the propagation equation of the velocity potential, ϕ, for an ideal fluid is

$$\nabla^2\phi - \frac{1}{c_1^2}\frac{\partial^2\phi}{\partial t^2} = 0, \tag{1}$$

where c_1 is the sound velocity. In the following, the subscript 1 refers to the fluid, and the subscript 2 to an inclusion.

Considering first the interaction between an incident wave and a single inclusion, three kinds of waves have to be considered: a compressional wave in the fluid, composed of the incident and of the scattered wave, respectively, ϕ_{inc} and ϕ_{c1}, and refracted compressional and shear waves in the solid inclusion, respectively, ϕ_{c2} and χ_{s2}. According to the symmetry, the velocity potentials of these waves can be expressed in terms of spherical Bessel functions for convergent waves and Hankel functions for divergent waves, and Legendre polynomials[10]

$$\phi_{\text{inc}} = \sum_{n=0}^{+\infty} i^n(2n+1)j_n(k_{c_1}r)P_n(\cos\theta), \tag{2}$$

$$\phi_{c_1} = \sum_{n=0}^{+\infty} i^n(2n+1)A_n h_n(k_{c_1}r)P_n(\cos\theta), \tag{3}$$

$$\phi_{c_2} = \sum_{n=0}^{+\infty} i^n(2n+1)B_n j_n(k_{c_2}r)P_n(\cos\theta), \tag{4}$$

$$\chi_{s_2} = -\sum_{n=1}^{+\infty} i^n(2n+1)C_n j_n(k_{s_2}r)\frac{\partial P_n}{\partial\theta}(\cos\theta), \tag{5}$$

where r and θ are the polar coordinates of the observing point in the system centered on the particle, j_n and h_n are the spherical Bessel and Hankel functions of nth order, P_n is the Legendre polynomial of the first kind of the nth order, k_c and k_s being the compressional and shear wave numbers, respectively. A_n, B_n, and C_n are weighting coefficients calculated from boundary conditions. This set of equations does not incorporate thermal waves in the fluid as was proposed by Allegra and Hawley.[11] These effects are supposed to be negligible in a solid–fluid medium. The A_n coefficients define the scattered longitudinal wave, so they are of main interest here. It can be noticed that the order of these developments is the multipole order which is a critical parameter in computation: the larger the scatterer is, the higher this order has to be.

These potentials verify a set of boundary conditions: continuity of the radial component of velocity, and of the radial and tangential components of the stress at the inclusion's surface

$$\frac{\partial}{\partial r}\left(\phi_{c_2} + \frac{\partial}{\partial r}(r\chi_{s_2})\right) + k_{s_2}{}^2 r\chi_{s_2}\bigg|_{r=a} = \frac{\partial}{\partial r}(\phi_{\text{inc}} + \phi_{c_1})\bigg|_{r=a}, \tag{6}$$

$$\mu\frac{\partial}{\partial\theta}\left(2\frac{\partial}{\partial r}\frac{1}{r}\left(\phi_{c_2} + \frac{\partial}{\partial r}(r\chi_{s_2})\right) + k_{s_2}{}^2 r\chi_{s_2}\right)\bigg|_{r=a} = \mu\frac{\partial}{\partial\theta}\left(2\frac{\partial}{\partial r}\frac{1}{r}(\phi_{\text{inc}} + \phi_{c_1})\right)\bigg|_{r=a}, \tag{7}$$

$$-\lambda k_{d_2}{}^2\phi_{c_2} + 2\mu\frac{\partial}{\partial r}\left(\frac{\partial}{\partial r}\left(\phi_{c_2} + \frac{\partial}{\partial r}(r\chi_{s_2})\right) + k_{s_2}{}^2 r\chi_{s_2}\right)\bigg|_{r=a} = -\lambda k_{d_1}{}^2(\phi_{\text{inc}} + \phi_{c_1}) + 2\mu\frac{\partial}{\partial r}\left(\frac{\partial}{\partial r}(\phi_{\text{inc}} + \phi_{c_1})\right)\bigg|_{r=a}, \tag{8}$$

where λ, μ are the Lamé coefficients, and a is the particle radius.

Combining these equations and Eqs. (2)–(5) leads to relations between coefficients A_n, B_n, and C_n which can be written in a 3×3 matrix form, known as the T-matrix

$$\begin{bmatrix} d_{11}^n & d_{21}^n & d_{31}^n \\ d_{12}^n & d_{22}^n & d_{32}^n \\ d_{13}^n & d_{23}^n & d_{33}^n \end{bmatrix} \cdot \begin{bmatrix} A_n \\ B_n \\ C_n \end{bmatrix} = \begin{bmatrix} b_1^n \\ b_2^n \\ b_3^n \end{bmatrix}, \tag{9}$$

where d_{ij}^n are coefficients and $b_j^n = \text{Re}(d_{ij}^n)$.[12,13]

This equation describes the interactions between a planar incident wave and a single spherical obstacle.[14]

The far-field scattering amplitude $f(\theta)$ defined as the ratio of the scattered wave to the incident wave, can now be calculated using the T-matrix

$$f(\theta) = \frac{i}{k_{c_1}} \cdot \sum_n (2n+1)\cdot P_n(\cos\theta)\cdot A_n. \tag{10}$$

For a set of scatterers distributed in a fluid matrix with an n_0 volume fraction, various scattering theories have been developed to predict the complex wave number in the effective medium. These theories express the coherent field, i.e., the average of the field over all configurations of scatterers. The wave number thus calculated describes the propagation after an averaging process.

The Waterman and Truell (WT) theory expresses the excitation field of the scatterer in position $\mathbf{r}_j$, noted $\phi(\mathbf{r}|\mathbf{r}_j)$, as the sum of the incident field and of the fields scattered by all other inclusions

$$\phi^E(\mathbf{r}|\mathbf{r}_j) = \phi_{\text{inc}}(\mathbf{r}) + \sum_{k\neq j}^{N} T(\mathbf{r}_k)\phi^E(\mathbf{r}|\mathbf{r}_k), \tag{11}$$

where $T(\mathbf{r}_k)$ represents the far-field amplitude radiated by a scatterer located in $\mathbf{r}_k$.

For a set of N scatterers defined by the vector of positions $\mathbf{r}_1\mathbf{r}_2\ldots,\mathbf{r}_N$ and by their scalar scattering properties $s_1,s_2,\ldots,s_N$, the configurational average can be defined as the average over all possible configurations, weighted by their distribution of probability. This average is over the positions, as well as over scattering properties; thus the average excitation field for a scatterer located in $\mathbf{r}_1$ is

$$\langle \phi^E(\mathbf{r}|\mathbf{r}_1)\rangle = \phi_{\text{inc}}(\mathbf{r}) + \int n(\mathbf{r}')T(\mathbf{r}')\langle \phi^E(\mathbf{r}|\mathbf{r}')\rangle - R(\mathbf{r}|\mathbf{r}_1), \tag{12}$$

where $\langle\ \rangle$ denotes the configurational average and $R(\mathbf{r}|\mathbf{r}_1)$ is a corrective term including spatial correlation between two scatterers located in $\mathbf{r}$ and $\mathbf{r}_1$. Here, these correlations are neglected. The configurational average of the excitation field $\langle \phi^E\rangle$ is supposed to be symmetric with the direction of propagation (z axis), so it can be developed as follows:

$$\langle \phi^E(\mathbf{r}|\mathbf{r}_1)\rangle = \sum_n E_n(z_1)\cdot j_n(k_{c_1}|r-r_1|)\cdot P_n(\cos\theta), \tag{13}$$

where E_n are the weighting coefficients.

Then, by expressing the incident wave (2) and the excitation field (13), Eq. (12) leads, by identification, through the extinction theorem (i.e., expression the extinction of the incident wave at the first interface of the scattering medium) to the expressions of the excitation coefficients E_n. The main result of the WT theory is then deduced: It is the expression of the wave number κ in the two-phase medium

$$\left(\frac{\kappa}{k_{c_1}}\right)^2 = \left[1+\frac{2\pi n_0 f(0)}{k_{c_1}^2}\right]^2 - \left[\frac{2\pi n_0 f(\pi)}{k_{c_1}^2}\right]^2, \tag{14}$$

where $f(0)$ and $f(\pi)$ are the far-field scattering amplitudes in forward and backward directions, calculated by (10). This wave number describes the coherent part of the scattered wave. It characterizes the mean effective medium.

Now, if statistical correlations in position are taken into account, the quasicrystalline approximation (QCA), first introduced by Lax, has to be considered. The average excitation field located in $\mathbf{r}$ taking into account two scatterers is close to the average excitation field with only one scatterer. Thus, the approximation used here is

$$\langle \phi^E(\mathbf{r}|\mathbf{r}_1,\mathbf{r}_2)\rangle = \langle \phi^E(\mathbf{r}|\mathbf{r}_1)\rangle. \tag{15}$$

Lax showed that the averaged field with one scatterer held fixed is related to the averaged field with two scatterers held fixed by

$$\langle \phi^E(\mathbf{r}|\mathbf{r}_1)\rangle = \phi_{\text{inc}}(\mathbf{r}) + \int n(\mathbf{r}_1|\mathbf{r}_2)T(\mathbf{r}_2)\langle \phi^E(\mathbf{r}|\mathbf{r}_1,\mathbf{r}_2)\rangle, \tag{16}$$

with $n(\mathbf{r}_1|\mathbf{r}_2)=n_0 p(|\mathbf{r}_1-\mathbf{r}_2|)$, where p is a pair-correlation function.

Statistical mechanics provides various pair-distribution functions for a set of spherical objects.[5] Here, we have chosen the solution of the Percus–Yevick (PY) equation because it is a good approximation of the correlation function between two hard spheres located in $\mathbf{r}_1$ and $\mathbf{r}_2$ and because it enables the calculation of the pair-correlation function up to 40% of volume fraction.[15] Figure 1 presents the pair-correlation function $p(\mathbf{r})$ used here, for various concentrations of particles with a diameter a.

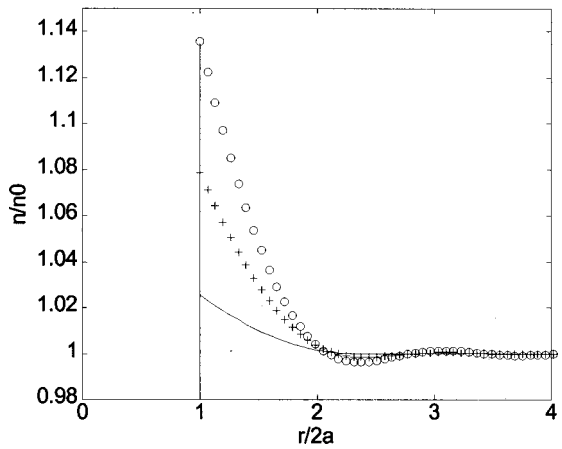


FIG. 1. Solutions of the Percus–Yevick equation for 1% (−), 5% (+), and 10% (○) volume fractions of hard spheres.

This pair-correlation function represents a hole (i.e., no scatterer) of radius a, as scatterers are impenetrable, and a spatially dependent concentration. The latter result means that the local volume fraction of scatterers can be higher or lower than the average volume fraction.

Expressing the fields of Eq. (16) in spherical Bessel functions leads to

$$A_n = 4\pi n_0 \sum_{m=0}^{+\infty}(2m+1)T_m A_m \sum_p \alpha(0,m|0,n|p) \times [L_p(k_{c_1},\kappa,a)+M_p(k_{c_1},\kappa,a)], \tag{17}$$

$$\kappa = k_{c_1} - \frac{2\pi i n_0}{k_{c_1}^2}\sum_{m=0}^{+\infty}(2m+1)T_m A_m, \tag{18}$$

with

$$L_p(k_{c_1},\kappa,a) = -\frac{a^2}{(\kappa^2-k_{c_1}^2)}[k_{c_1}h_p'(k_{c_1}a)j_p(\kappa a) - \kappa h_p(k_{c_1}a)j_p'(\kappa a)], \tag{19}$$

$$M_p(k_{c_1},\kappa,a) = \int_a^{+\infty} r^2(p(r)-1)h_p(k_{c_1}r)j_p(\kappa r)dr, \tag{20}$$

where $\alpha(0,m|0,m|p)$ is the Wigner $3j$ symbol, the T_l are the far-field scattering coefficients, calculated from the T-matrix, and a is the particle diameter corresponding to the shortest distance between two centers of scatterers.

Spatial correlations are taken into account by the L_p factor (correction for impenetrable scatterers, known as the hole correction[7]), and M_p factor (correction for the PY pair-correlation function). Now, the effective wave number κ is determined by the secular equation of system (17).

TABLE I. Properties of ethylene–glycol and acrylic.

	Density	Cc (m/s)	Cs (m/s)
Ethylene–glycol (25 °C)	1.113	1658	···
Acrylic	1.18	2610	1100

III. EXPERIMENTAL SETUP

Attenuation measurements were performed by a transmission method at constant thickness. The phase velocity and the attenuation are measured by comparison with a reference signal in a nonattenuating medium. The diffraction effects have been estimated by calculating the average particle displacement over the radius of the receiving transducer.[16] This leads to less than 1% of diffraction effect in attenuation and less than 2×10^{-3} rad in phase. Diffraction will therefore be neglected. Furthermore, acoustic impedance corrections on the electroacoustic responses of the transducers were evaluated and found to be negligible. The received signal is digitized on 10.5 bits at 50 MHz by a Lecroy 9430 oscilloscope and averaged over 5000 acquisitions. The data are then transmitted by a GPIB protocol to a personal computer to be analyzed.

Broadband transducers of center frequencies 5 and 10 MHz excited by a short electrical pulse were used in order to obtain attenuation and phase velocity measurements from 1.5 to 18 MHz. The distance between the transducers, around 20 mm, was measured precisely by determining the time of flight separating the first and second echoes. Their shapes are close enough so that cross correlation can be used in ethylene–glycol at ambient temperature.

The samples investigated here are composed of ethylene glycol in which acrylic (PMMA) beads are in suspension, the properties of which are listed in Table I. Two particle-size distributions, the characteristics of which were obtained by a laser technique, were used. The first, referred to as group A, has a 120-μm average diameter with a standard deviation of 30 μm, and the second, referred to as group B, has a 215-μm average diameter with a standard deviation of 60μm. Suspensions with volume fractions of scatterers between 1% and 10% were investigated. Samples were homogenized constantly by a magnetic agitator to obtain many geometrical configurations. The averaging process over a long period of time in a moving medium enables the measurement of the coherent effects of multiple scattering. Experiments began with pure ethylene–glycol; scatterers were then added with 1% increment steps.

IV. ATTENUATION AND VELOCITY MEASUREMENTS

Attenuation is the difference between the amplitude spectrum (in dB) of the signal in ethylene–glycol (reference signal) and the amplitude spectrum of the signal in the sample. The result is then divided by the sample thickness.

It is important to pay attention to the phase velocity calculation. Indeed, phase velocity is determined from the unwrapped phase spectrum of the reference signal and the signal in the sample. Most of the unwrapping processes begin the computation with the very first point of the phase spectrum. This point is often out of the passband of the transducers, so the measurement is not reliable. In order to eliminate the vertical shift due to this error, the unwrapping process must be initiated in the passband and before the first discontinuity of the wrapped (acquired) phase. It is therefore necessary to shift the signals at the beginning of the acquisition window, in order to have a continuous phase spectrum on a frequency domain as wide as possible. The starting point of the unwrapping process is then chosen in this domain.

V. RESULTS AND DISCUSSION

A. Simulation results

The WT and the quasicrystalline approximation using the Percus–Yevick equation (QCA-PY) models have been implemented. Both of them use the T-matrix formalism with compressional and shear waves in inclusions, and only compressional waves in the surrounding medium.

First, the effect of the particle-size distribution is studied [scatterers are assumed to be spherical with radius varying with a normal distribution $\sigma(a)$]. To take into account this effect in the WT model, as proposed by Waterman, a new far-field coefficient is calculated assuming that it is a sum of the far-field coefficients of each group of scatterers weighted by their granulometric distribution.

$$\langle f(\theta)\rangle = \int_0^\infty \sigma(a) f_a(\theta)\, da. \tag{21}$$

In the QCA-PY model, the granulometric dispersion cannot be taken into account in the far-field coefficients because they are not explicit. Thus, an average T-matrix has to be defined. The coefficients d_{ij}^n and b_j^n in the T-matrix are now average coefficients calculated from each group of scatterers weighted by their granulometric distribution.

$$\langle d_{ij}^n\rangle = \int_0^\infty \sigma(a) d_{ij}^n(a)\, da. \tag{22}$$

In this model, the chosen pair-correlation function is the solution of the Percus–Yevick equation for hard particles with a diameter equal to the mean diameter $\langle a\rangle$ of the distribution. Such an assumption can be made here because the diameter distribution is narrow around $\langle a\rangle$.

Figure 2 shows the attenuation for identical 90-μm particles and a normal distribution of diameters with a mean value $\langle a\rangle$ of 90 μm and a standard deviation of 20 μm, while Fig. 3 compares attenuation for normal distributions with a mean value of 90 μm and standard deviations of 10 and 20 μm, both with the WT theory.

When scatterers are assumed to be identical, sharp peaks due to resonances in the scatterers are observed. The positions of these peaks are related to the size of inclusions. When dispersion is introduced, the contribution of each size becomes very small: the averaging process leads to the extinction of the peaks. When the dispersion increases, the maximum of attenuation is damped and shifted to low normalized frequencies. Indeed, scattering is dominated by the contribution of large particles. When the normalized fre-

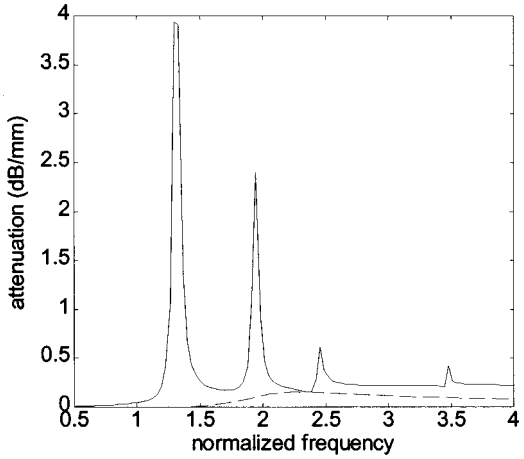


FIG. 2. Attenuation in a 5% volume fraction suspension of acrylic scatterers in ethylene–glycol, predicted by WT for a single size of 90 μm (solid line) and for a normal distribution of diameters with a mean value of 90 μm and a standard deviation of 20 μm (dashed line).

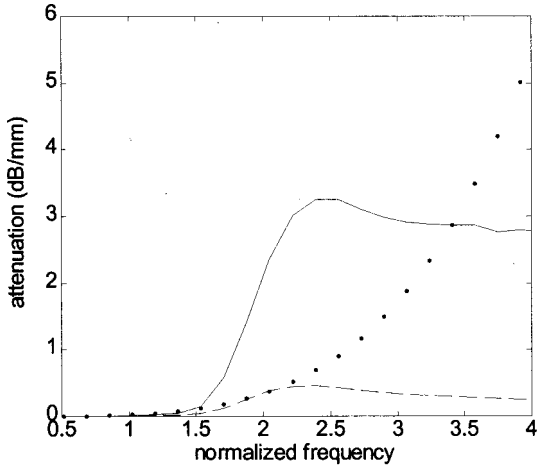


FIG. 4. Attenuation versus normalized frequency for a suspension of 3% of acrylic scatterers in ethylene–glycol with a mean particle diameter of 90 μm predicted by WT (solid line), QCA-PY (long dashed line) and Rayleigh (short dashed line).

quency is around unity, the behavior of attenuation with frequency is strongly influenced by the granulometry of suspensions.

Attenuation versus frequency curves have been calculated using the WT and QCA-PY models. Their predictions are presented in Fig. 4 for a suspension of 3% volume fraction of scatterers with a mean diameter of 90 μm. The results of the Rayleigh theory in the same situation, but without size distribution, are also plotted.

Both multiple-scattering models agree on the general behavior of attenuation. They predict a low attenuation for low frequency, very close to that predicted by Rayleigh scattering theory. When the normalized frequency increases, the attenuation drastically increases to reach a local maximum, and then remains almost constant. The mean value of scatterer diameters determines the position of the ''cutoff'' frequency because it represents the frequency of the first resonance of the main part of the population of scatterers. The width of the peak of attenuation is related to the standard deviation of the size distribution. Finally, it can be seen that the two models differ in the prediction of attenuation: the QCA-PY model predicts an attenuation around five times higher than that of the WT theory. This is due to the fact that with WT theory, the interaction between two scatterers is not taken into account.

Figure 5 presents the prediction of the WT and QCA-PY models for the phase velocity versus frequency in a suspension of 3% volume fraction of scatterers with a mean diameter of 90 μm.

Phase velocity prediction of these two models is very close for low volume fractions. At low frequencies, both models predict a velocity of 1661.8 m/s, which is very close to that predicted by the Rayleigh single-scattering theory. When the normalized frequency is lower than unity, the Rayleigh theory remains valid. However, it does not predict the frequency behavior of velocity. When the frequency increases, velocity decreases to a minimum at the frequency corresponding to the cutoff observed in attenuation, then the velocity rapidly increases and finally remains constant for high normalized frequencies. The velocity dispersion predicted by the WT and QCA-PY models is very close when the normalized frequency is lower than 2. Differences be-

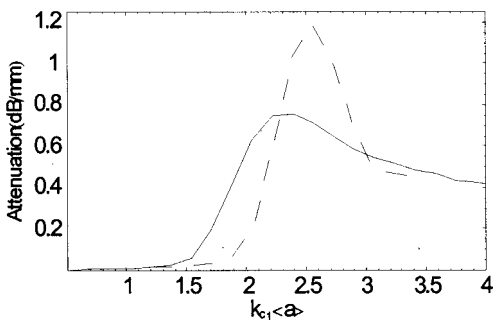


FIG. 3. Attenuation versus normalized frequency in a 5% volume fraction suspension of acrylic scatterers in ethylene–glycol, predicted by WT for a normal distribution of diameters with a mean value of 90 μm and standard deviations of 10 μm (dashed line) and 20 μm (solid line).

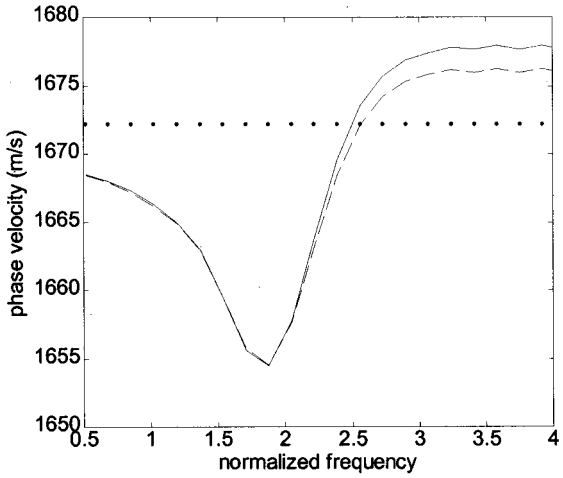


FIG. 5. Phase velocity versus normalized frequency in a suspension of 3% of acrylic scatterers in ethylene–glycol with a mean particle diameter of 90 μm predicted by WT (solid line), QCA-PY (dashed line).

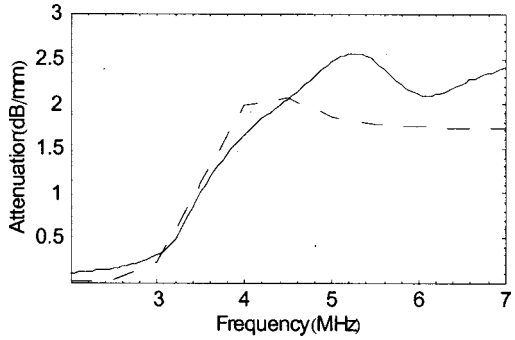


FIG. 6. Attenuation versus frequency in a 3% volume fraction suspension of acrylic spheres in ethylene glycol. Theoretical values by the QCA-PY model with a mean particle diameter of 150 μm and a standard deviation of 30 μm (dashed line), and experimental results (solid line).

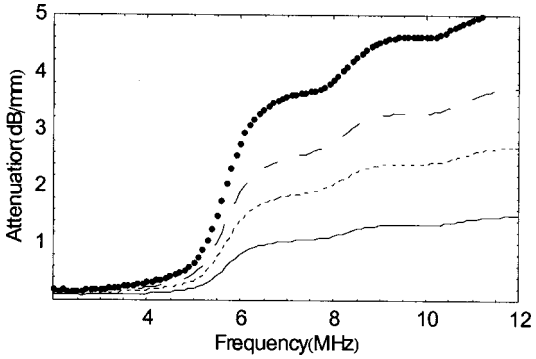


FIG. 8. Attenuation measurements versus normalized frequency in a suspension of acrylic spheres in ethylene–glycol at 1% (solid line), 2% (short dashed line), 3% (long dashed line), and 4% (points) volume fractions.

come greater as normalized frequency increases. The WT model prediction is lower than the QCA-PY one, as spatial correlations are not included.

It appears that spatial correlation is a critical parameter in the description of acoustic scattering in such suspensions, especially to predict attenuation. Even at low concentrations, neglecting these correlations in suspensions leads to an underestimation of attenuation. Phase velocity is also influenced by these correlations but only at high volume fractions or normalized frequency.

B. Experimental results and discussion

Attenuation and phase velocity have been measured versus frequency for suspensions with volume fractions of 1%, 2%, and 3% for the two groups of scatterers.

Figures 6 and 7 present theoretical attenuation and phase velocity obtained with the QCA-PY model and experimental results.

In these calculations, the granulometric parameters have been adjusted in order to fit the experimental data. Figures 6 and 7 have been obtained for a normal distribution with a mean diameter of 150 μm and a standard deviation of 30μm. The optical characterization results were 215 μm for the mean diameter and 60 μm for the standard deviation. This laser technique is based on a volume measurement of particles, which means that large particles can be overestimated while small ones may be not detected. Again, the attenuation is underestimated at low frequencies because of the viscous phenomena. The position of the calculated maximum is lower than the experimental and its amplitude is also lower. Between 5 and 9 MHz, the predicted attenuation remains constant, but in measured data, variations are observed. All these differences are probably due to the effective size distribution, which is not perfectly Gaussian. The first resonance in small scatterers or the second resonance in larger ones may be responsible for the observed oscillations. Furthermore, it appears that the viscosity losses, which are not included in the model, cannot be neglected. The theoretical and experimental behaviors of the phase velocity agree very well. The only discrepancy that can be seen is due to a difference of temperature (25 °C for the modeled phase velocity and around 24 °C during the experiment), which induces a vertical shift of a few meters per second.

Figures 8 and 9 show attenuations and phase velocities measured at different volume fractions.

The behaviors of attenuation and velocity are similar at the different volume fractions, but the amplitude of the variations increase with volume fraction: it is experimentally verified that attenuation increases with concentration, as scattering dominates viscous effects.

Three important parameters of suspensions can be evaluated with the phase velocity curve. The position of the minimum and the width of this peak are related to size-

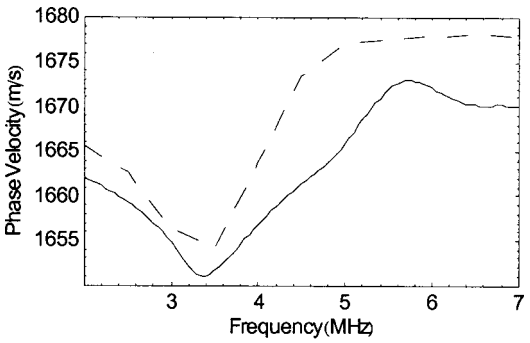


FIG. 7. Phase velocity versus frequency in a 3% volume fraction suspension of acrylic spheres in ethylene–glycol. Theoretical values by the QCA-PY model with a mean particle diameter of 150 μm and a standard deviation of 30 μm (dashed line), and experimental results (solid line).

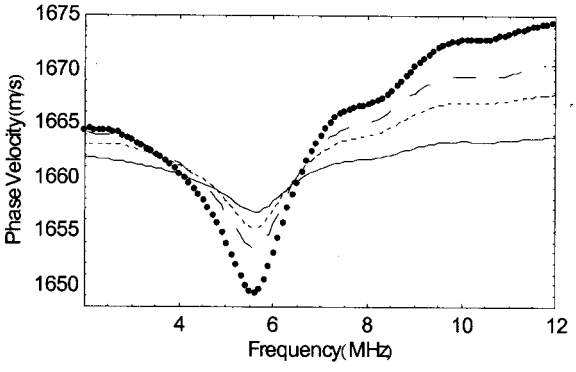


FIG. 9. Phase velocity measurements versus normalized frequency in a suspension of acrylic spheres in ethylene–glycol at 1% (solid line), 2% (short dashed line), 3% (long dashed line), and 4% (points) volume fractions.

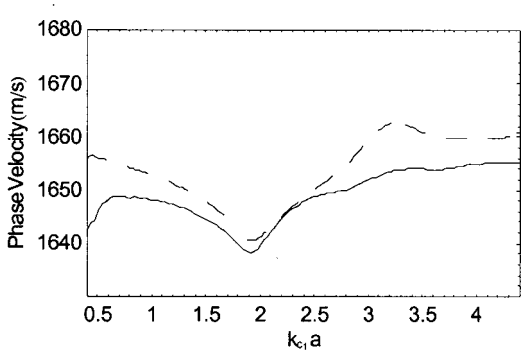


FIG. 10. Experimental phase velocity versus normalized frequency in a 3% volume fraction suspension of acrylic spheres in ethylene–glycol, with a mean particle diameter of 90 μm and a standard deviation of 20 μm (dashed line) and with a mean particle diameter of 150 μm and a standard deviation of 30 μm (solid line).

distribution parameters, i.e., the mean diameter of particles and the standard deviation. The amplitude of velocity variation is related to the volume fraction of the suspension.

Figures 10 and 11 represent experimental velocity and attenuation versus normalized frequency in suspensions of 3% volume fraction of acrylic scatterers in ethylene–glycol.

As expected, velocity and attenuation behaviors as a function of normalized frequency are the same for the two groups of scatterers. The discrepancies observed at low frequencies are due to the viscous attenuation in the fluid matrix (the absolute frequencies are not the same for the two groups of scatterers). At high frequencies, differences observed are due to viscous interactions between fluid and particles. Differences in the standard deviation of the particle-size distributions could also be the cause of some discrepancies.

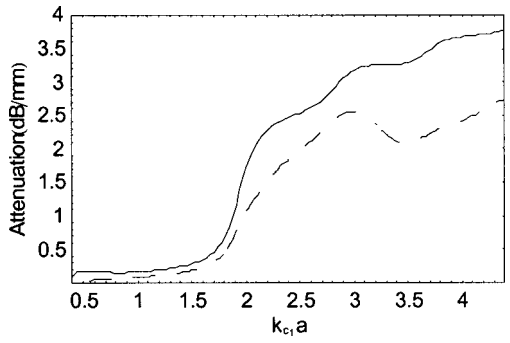


FIG. 11. Experimental attenuation versus normalized frequency in a 3% volume fraction suspension of acrylic spheres in ethylene–glycol, with a mean particle diameter of 90 μm and a standard deviation of 20 μm (dashed line) and with a mean particle diameter of 150 μm and a standard deviation of 30 μm (solid line).

VI. CONCLUSION

The behaviors of attenuation and phase velocity have been theoretically and experimentally studied for suspensions of solid particles in a fluid matrix, in the case where the particle size is close to the wavelength, and for volume fractions up to 8%. Experimental results were compared to theoretical results of two multiple-scattering theories in which we have included the size distribution of scatterers. This size distribution is a critical parameter to understand the frequency behavior of phase velocity and attenuation.

Furthermore, results show the major role in the prediction of attenuation of spatial correlation between the position of scatterers.

Future work will concentrate on including viscothermal effects in theoretical models and developing inverse procedure to determine granulometric properties from phase velocity and attenuation measurements.

[1] Lord Rayleigh, *The Theory of Sound* (Dover, New York, 1945).
[2] J. R. Allegra and S. A. Hawley, ''Attenuation of sound in suspensions and emulsions: Theory and experiments,'' J. Acoust. Soc. Am. **51**, 1545–1563 (1972).
[3] V. Twersky, ''On scattering of waves by random distributions,'' J. Math. Phys. **3**, 724–734 (1962).
[4] L. L. Foldy, ''The multiple scattering of waves,'' Phys. Rev. **67**, 107–119 (1945).
[5] P. C. Waterman and R. Truell, ''Multiple scattering of waves,'' J. Math. Phys. **2**(4), 512–537 (1961).
[6] M. Lax, ''Multiple scattering of waves. The effective field in dense systems,'' Phys. Rev. **85**(4), 621–629 (1952).
[7] J. G. Fikioris and P. C. Waterman, ''Multiple scattering of waves. Hole corrections in the scalar case,'' J. Math. Phys. **5**(10), 1413–1420 (1964).
[8] L. Tsang, J. A. Kong, and T. Habashy, ''Multiple scattering of acoustic waves by random distribution of discrete spherical scatterers with the quasi-crystalline and Percus–Yevick approximation,'' J. Acoust. Soc. Am. **71**, 552–558 (1982).
[9] Y. Ma, V. V. Varadan, and V. K. Varadan, ''Multiple scattering theory for wave propagation in discrete random media,'' Int. J. Eng. Sci. **22**, 1139–1148 (1984).
[10] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1986).
[11] J. R. Allegra and S. A. Hawley, ''Attenuation of sound in suspensions and emulsions: Theory and experiments,'' J. Acoust. Soc. Am. **51**, 1545–1564 (1971).
[12] G. C. Gaunaurd and W. Wertman, ''Comparison of effective medium theories for inhomogeneous continua,'' J. Acoust. Soc. Am. **85**, 541–554 (1989).
[13] A. M. Baird, F. H. Kerr, and D. J. Townsend, ''Wave propagation in a viscoelastic medium containing fluid-filled microspheres,'' J. Acoust. Soc. Am. **105**, 1527–1538 (1999).
[14] N. D. Veksler, *Resonance Acoustic Spectroscopy* (Springer, Berlin, 1993).
[15] J. K. Percus and G. J. Yevick, ''Analysis of classical statistical mechanics by means of collective coordinates,'' Phys. Rev. **110**(1), 1–13 (1958).
[16] G. S. Kino, *Acoustic Waves* (Prentice-Hall, Englewood Cliffs, NJ, 1987), pp. 171–175.

Acoustic flowmeter for the measurement of the mean flow velocity in pipes

Wan-Sup Cheung,[a] Hyu-Sang Kwon, Kyung-Am Park, and Jong-Seung Paik
Mechanical Metrology Division, Korea Research Institute of Standards and Science, P.O. Box 102, Yusong, Taejon 305-600, Republic of Korea

(Received 5 June 2000; revised 25 July 2001; accepted 27 July 2001)

This paper proposes a new technique for measuring the gas flow velocity averaged along the finite length of a pipe as well as over its cross-sectional area. Unlike the conventional gas flowmeters, the proposed technique exploits the one-dimensional plane waves that propagate uniformly across the pipe cross-sectional area. When a fluid flows along the pipe, the plane waves are superposed with the flow field such that the positive-going and negative-going plane wave components undergo the change of their wave numbers. Such wave number variation due to the mean flow velocity has provided a major motivation for developing a new way of measuring the mean flow velocity in the pipe, which is referred to as the acoustic flowmeter. To examine the feasibility of the developed flow velocity measurement method, including its theoretical backgrounds, experimental setups are illustrated in this paper. Detailed experimental data for the flow velocity range of 2–27 m/s reveal the linearity of the proposed acoustic flowmeter and its salient environmental robustness for the different acoustic pressure patterns in the pipe and, furthermore, for different velocity profiles over the pipe cross-section area. [DOI: 10.1121/1.1404379]

PACS numbers: 43.20.Ye, 43.28.We [LCS]

I. INTRODUCTION

The measurement technology of gas flow velocity in pipes has recently been an important issue in nationwide natural gas pipelines, intake and exhaust gas pipelines of automobiles, and ventilation ducts of large buildings. The gas flow measurement has been a prime factor in the state control of gas supply systems, for the production of quieter and more fuel-efficient engines and in the environmental noise reduction of large buildings. The flow measurement principle in these industrial applications is still based on the classical nozzle-type or hot-wire flowmeters.[1] Unlike these classical flow measurement methods, acoustical methods have attempted to use the principle of the pipe or duct acoustics[2] to measure the gas flow velocity. Fundamentals behind these attempts are based on the use of the incident and reflected plain waves in the pipe. Potzick *et al.*[3] suggested the long wavelength acoustic flowmeter that uses the two microphones installed in the pipe wall to measure the standing waves excited by two frequency components (fundamental and second-harmonic components). The specific equal-amplitude conditions for the two microphones, which are obtained by choosing the two harmonic components of a sound source, were shown to give a way of measuring the flow velocity in the pipe. This method requires a delicate tracking and tuning technology of two harmonic sounds that should satisfy the equal amplitude conditions according to the flow velocity. The instrumentation for tracking and tuning two harmonic sounds presents much difficulty in real implementation. Unlike this tracking and tuning methodology, Alves *et al.*[4,5] suggested the use of the acoustical transfer function between paired microphones spaced a finite distance apart. The phase change of the transfer function, which depends on the flow velocity and its direction, was exploited to estimate the Mach number of flowing fluid in the pipe. Three different methods of estimating the phase change of the standing wave and related Mach number using the single, dual, and triple paired microphones were introduced recently. Feasibility studies on these methods are in progress, including experimental investigation to judge their effectiveness in real flow measurements.

This work is also based on fundamental duct acoustics.[2] Acoustic plane waves in the pipe present two excellent properties: the uniformly distributed propagation velocity normal to the cross-section area and the outstanding repeatability of the plane waves inside the pipe. The uniformly distributed sound propagation property leads to the measurement of the averaged flow velocity over the cross-sectional area. Furthermore, multiple velocities at multiple positions, locally measured within the finite length, enable us to obtain the mean velocity along the length of the pipe as well as over the pipe cross-section area. In fact, the acoustic plane waves in the pipe are always accompanied by the superposition of the ''incident'' plane wave coming from the acoustic source with the reflected one from the outside of the straight pipe. This condition provides excellent repeatability of the acoustic plane wave inside the pipe. The ''effective'' speed of sound for the plane wave propagating in the same direction as fluid flow is increased by the amount of the normal flow velocity. Similarly, the ''effective'' speed of sound for the plane wave propagating oppositely to the flow direction is decreased. Both changes of wave numbers corresponding to each plane wave can give an indirect measure of the averaged flow velocity along the finite length of the pipe as well as over the cross-sectional area. This understanding has already been investigated in experimental work.[6–8] More improved mea-

[a]Electronic mail: wansup@kirss.re.kr

surement techniques of flow velocity in the pipe are proposed in Sec. II. In order to examine the proposed methods experimentally, a new measurement device developed in this work is also introduced in Sec. III. The experimental reconstruction of acoustic pressure field in the measurement system is addressed in Sec. IV. Experimental results are presented to demonstrate the success of this work and related discussions are also added in Sec. V. In Sec. VI, major contributions of this work are summarized and work in progress is discussed.

II. THEORETICAL BACKGROUNDS AND MEASUREMENT TECHNIQUES

A general case of sound propagation in ducts[2] may include the convective and scattering effects of the mean flow with refraction by velocity gradients and temperature gradients. But plane wave analysis is appropriate for many cases of sound propagation in ducts where the wavelength of interest is large compared to their transverse dimension. A good example is the fluid conveying pipelines where flow measurements are needed.

A. Acoustic plane wave propagation in the pipes

When an acoustic wave, whose wavelength λ is quite large compared to the inner diameter D of the pipe (i.e., typically $\lambda > 2D$ as recommended in ISO 10534-1),[9] is generated in the pipe, it becomes a plane wave whose acoustic pressure distribution is very nearly uniform[10] over the cross section after it is propagated more than a wavelength from the source. When the mean flow velocity field is superposed with the acoustic field in the pipe, the resultant acoustic pressure is described as

$$P(x,t) = P(x)P(t) = (P_0^+ e^{-jk^+ x} + P_0^- e^{jk^- x}) e^{i\omega t}. \tag{1}$$

P_0^+ and P_0^- are the acoustic pressures of positively and negatively propagating wave components at position $x=0$. ω denotes the angular velocity of sound ($\omega = 2\pi f$, f = the excitation frequency of an acoustic source) and k^+ and k^- are the complex wave numbers corresponding to the positively and negatively propagating plane waves. The complex wave numbers are defined as

$$k^+ = (1-ja)k_0/(1+M), \quad k^- = (1-ja)k_0/(1-M), \tag{2}$$

where M is the Mach number ($M = U/C$), α the acoustic pressure attenuation constant,[2] and k_0 denotes the wave number of the stationary fluid ($U=0$). In this work, the attenuation constant is ignored because α is close to 0.0014 in air ($f = 500$ Hz, $D = 0.1$ m, $C = 343.3$ m/s). When the fluid in the pipe flows in the positive x direction, Eq. (2) indicates that the wave number of the positively propagating plane wave component decreases but that of the negatively propagating one oppositely increases. It means that the change in wave number for both plane wave components due to the flow velocity can be obtained quantitatively from the variation of the pressure pattern along the pipe. It has been a key idea in developing a new method to measure the mean flow velocity.

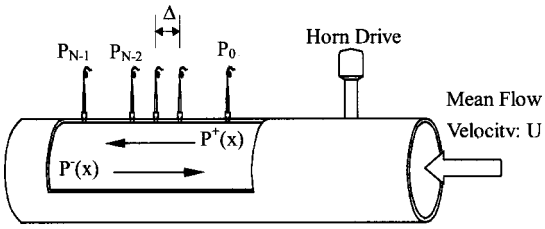


FIG. 1. Schematic setup of measuring the standing wave in the pipe.

B. Measurement method for the mean flow velocity in the pipe

From Eq. (1), the periodic standing pressure pattern along the pipe is separated into the spatially dependent term $P(x)$ and the time-dependent $P(t)$. Thus, the position-dependent acoustic pressure $P(x)$ in the pipe at any instant is considered separately. Figure 1 shows the schematic view of measuring the acoustic pressures of the standing wave pattern using multiple microphones installed on the pipe wall. When each microphone is equally spaced by a gap Δ, the acoustic pressures at three neighboring positions satisfy the following relations:

$$\begin{bmatrix} P_{n-1} \\ P_n \\ P_{n+1} \end{bmatrix} = \begin{bmatrix} e^{-jk^+(n-1)\Delta} & e^{jk^-(n-1)\Delta} \\ e^{-jk^+ n\Delta} & e^{jk^- n\Delta} \\ e^{-jk^+(n+1)\Delta} & e^{jk^-(n+1)\Delta} \end{bmatrix} \begin{bmatrix} P_0^+ \\ P_0^- \end{bmatrix},$$

$$n = 1,2,3,\ldots,N-2\,. \tag{3}$$

In Eq. (3), P_n denotes the acoustic pressure at position $x = n\Delta$. By solving the incident and reflected waves $\{P_0^+, P_0^-\}$ at $x=0$ from the first two acoustic pressures $\{P_{n-1}, P_n\}$ and then substituting the incident and reflected waves into the third pressure P_{n+1}, a spatial recursive from along the longitudinal position is obtained as

$$P_{n+1} e^{jk^+\Delta} + P_{n-1} e^{jk^-\Delta} = P_n (1 + e^{j(k^+ + k^-)\Delta}). \tag{4}$$

This spatial recursive form will be shown to play a central role in estimating the flow velocity in this paper.

The range of Mach number considered in this work is equal to or less than 0.1 (mean air flow velocity $U \leqslant 34.3$ m/s). Consequently, the positive and negative wave numbers are approximated as $k^+ \cong k_0(1 - M + M^2 - M^3)$ and $k^- \cong k_0(1 + M + M^2 + M^3)$ by neglecting the fourth- and higher-order terms of the Mach number. Similarly, the phase angle of the right-hand side of Eq. (4) is also approximated as $(k^+ + k^-)\Delta \cong 2k_0\Delta(1 + M^2)$. By substituting these approximations into the spatial recursive equation (4) and then arranging them, the spatial recursive relation is rewritten as

$$P_{n+1} e^{-j(M+M^3)k_0\Delta} + P_{n-1} e^{j(M+M^3)k_0\Delta} = 2\cos((1+M^2)k_0\Delta) P_n\,. \tag{5}$$

This presents a clear graphical understanding about the estimation of the Mach number using the three consecutive acoustic pressures $\{P_{n-1}, P_n, P_{n+1}\}$. Let the Mach number-dependent rotation angle be $\theta(M) = k_0\Delta(M + M^3)$. Then, the sum of the $(n+1)$th acoustic pressure P_{n+1} rotated negatively by $\theta(M)$ and the $(n-1)$th acoustic pressure

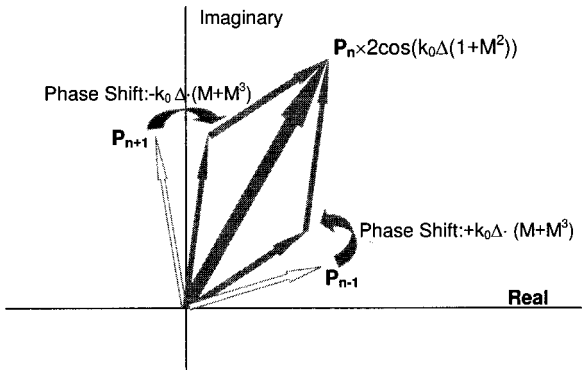


FIG. 2. Phase vector diagram of the spatial recursive form for three neighboring pressures.

P_{n-1} rotated positively by $\theta(M)$ is found to be equal to the nth acoustic pressure multiplied by $2\cos((1+M^2)k_0\Delta)$. Figure 2 illustrates this relation graphically. It shows that the Mach number is related to the amount of phase shift for the neighboring phase vectors P_{n+1} and P_{n-1} with respect to the midposition phase vector P_n. As the Mach number increases, the amount of the phase shift gets larger. Of course, the direction of the flow velocity is shown to determine the direction of the phase shift for the neighboring phase vectors by checking whether its direction is clockwise or counter clockwise. The rate of the phase shift is proportional to the gain factor $k_0\Delta$. This means that one can adequately choose the gain factor by varying either the excitation frequency of an acoustic source ($k_0=2\pi f/C$) or the gap Δ between the neighboring microphones. As a result, the sensitivity of the phase shift for the given Mach number can be suitably tuned in a real experiment.

C. Estimation of flow velocity

This section presents a way of estimating the best-fitted Mach number for multiple acoustic pressure measurement data along the pipe. As shown in Eq. (5) and Fig. 2, the Mach number-dependent rotation angle is described as $\theta(M)=k_0\Delta(M+M^3)$. Let the angle-related cosine and sine components be $C_M=A_M\cos(k_0\Delta(M+M^3))$ and $S_M=A_M\sin(k_0\Delta(M+M^3))$ where a scale factor $A_M=0.5/\cos(k_0\Delta(1+M^2))$ from Eq. (5). From the set of N measurements $\{P_n, n=0,1,2,...,N-1\}$, we obtain $N-2$ spatial recursive forms between three neighboring data, which are described as

$$\begin{bmatrix} (P_2+P_0) & j(P_0-P_2) \\ (P_3+P_1) & j(P_1-P_3) \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ (P_{N-1}+P_{N-3}) & j(P_{N-3}-P_{N-1}) \end{bmatrix} \begin{bmatrix} C_M \\ S_M \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \\ \cdot \\ \cdot \\ \cdot \\ P_{N-2} \end{bmatrix}. \tag{6}$$

For the case of $N \geqslant 4$, the best fitted coefficient vector $\mathbf{x}=[C_M, S_M]^T$ is readily obtained using the well-known least-squares method. Let the $(N-2)$-by-2 matrix at the left-hand side of Eq. (6) be $\mathbf{A}$ and the column vector at the left-hand side by $\mathbf{y}$. Then, the best-fitted coefficient vector $\mathbf{x}$ is always obtained from the following normal equation:

$$\begin{bmatrix} C_M \\ S_M \end{bmatrix} = [\mathbf{A}^H\mathbf{A}]^+\mathbf{A}^H\mathbf{y}, \tag{7}$$

where the superscript H denotes the Hermitian operator and $[\cdot]^+$ the pseudoinverse.[11] In this work, the singular value decomposition method presented in MATLAB[12] is used to obtain the pseudoinverse. Given the best-fitted coefficient vector $\mathbf{x}=[C_M, S_M]^T$, the Mach number is calculated from

$$M^3+M=\frac{1}{k_0\Delta}\tan^{-1}\left(\frac{S_M}{C_M}\right). \tag{8}$$

This third-order cubic equation gives a single real-valued Mach number that is readily calculated from the algebraic solution.[13] This real-valued solution is used in this work but two complex-valued ones are not considered because they do not have physical interpretation. It is obvious that the mean flow velocity obtained from the above-given estimation procedures is the best-fitted value for the locally measured acoustic pressure data that are collected over the pipe length between the first and final microphones. It should be noted from Eq. (8) that the ratio of the cosine component C_M to the sine component S_M enables the Mach number estimation without any knowledge of the Mach number-dependent scale factor $A_M=0.5/\cos(k_0\Delta(1+M^2))$.

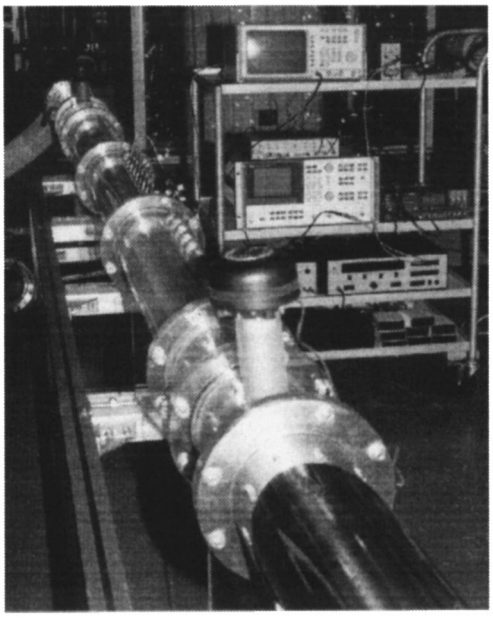


FIG. 3. Picture of experimental setup.

III. MEASUREMENT SETUPS

In order to examine experimentally the mean flow velocity measurement method introduced in Sec. II, experimental setups in Fig. 3 were built up. They consisted of two flow velocity measurement setups using the nozzle-type reference flowmeter and the acoustic flowmeter developed in this work.

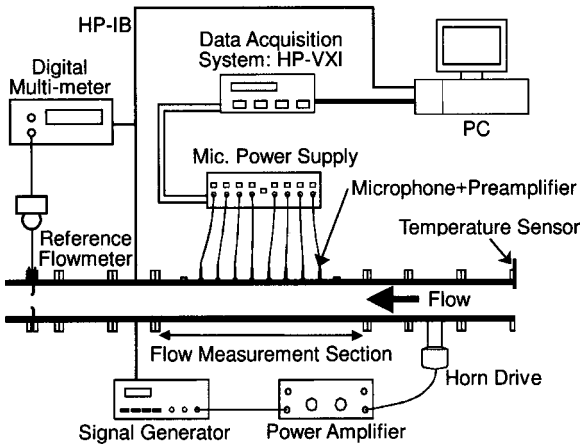


FIG. 4. Schematic configuration of measuring instruments.

A. Measurement setup of reference flowmeter

The schematic drawing of the experimental setups is shown in Fig. 4. In order to suck air into the pipeline, a blower was used. The flow rate of Mach number $M<0.1$ was controlled with a main butterfly valve and a bypass valve installed at the end of the downstream pipe. The reference nozzles were used to measure the mean air flow velocity in the test line of pipe diameter 100 mm. Three nozzles (throat diameter 23.4, 35.1, and 52.6 mm) with a differential pressure gauge were calibrated in the gas flow standard system of the Korea Research Institute of Standards and Science (KRISS). The calibrated uncertainty of mean velocity measurement for the three nozzles was found to be less than $\pm 0.43\%$ with a confidence level of 95%. The upstream straight length of a nozzle is 20 times the pipe diameter, which is enough upstream length for fully developed flow at a nozzle. To measure the mean air flow velocity, the upstream pressure was measured using a U-type manometer and the differential pressure of the nozzle was also measured using the pressure gauge (Rosemount Measurement Modal 1151). The output of the pressure gauge was read each half second through the digital multimeter (HP 34401A) that is interfaced to a PC using the HP–IB interface.

B. Measurement setup of acoustic flowmeter

The instruments shown in Fig. 4 were used to measure the flow velocity using the acoustic wave. The temperature sensor was used to calculate the speed of sound flowing in the pipe. The horn-type acoustic source that generates sound in the pipe was installed at the upstream part of the flow measurement section whose inner diameter is 0.1 m. In order to generate a pure tone sound in the pipe, the signal generator (HP 33120A with a phase lock/TCXO time base) was chosen, which is interfaced to the PC using the HP–IB interface. Its output is connected to the power amplifier (B&K 2706) that provides an adequate current level for the horn driver unit. A microphone array, which consists of seven equispaced $\frac{1}{4}$ in. microphones (GRAS 2633) and their preamplifiers (GRAS 26AC), was also fixed just on the inner wall of the measurement section. The gap between two microphones was chosen to be 0.1 m ($\Delta = 0.1$ m). The multichannel microphone power supply (Larson–Davis model 2210) was used to operate the seven-channel microphones and their preamplifiers. The seven-channel acoustic pressure signals were simultaneously converted into a digital time series by using the HP-VXI data acquisition system (dual HP E8402 A's) that simultaneously converts all analog signals into 16 bit digital codes and records them digitally into the PC hard drive.

IV. RECONSTRUCTION OF PHASE VECTORS

This section presents a way of estimating the phase vector of acoustic pressure at each acoustic pressure measurement position from the acoustic pressure signals sampled in the time domain. Let $\{p_m(x_n,k\Delta t);\ n=0,1,...,N-1$ and $k=0,1,2,...\}$ be the set of digitally sampled acoustic pressure signals in time domain (Δt=sampling time, N=number of microphones, position $x_n=n\Delta$). Given the single frequency f of the acoustic source, it is quite straightforward to estimate the phase vectors at the measurement positions by calculating the discrete Fourier series of the sampled pressure data, as follows:

$$C_n=C(x_n,f)=\frac{2}{K_0}\sum_{k=0}^{K_0-1} p_m(x_n,k\Delta t)\cos(2\pi fk\Delta t),$$
$$S_n=S(x_n,f)=\frac{2}{K_0}\sum_{k=0}^{K_0-1} p_m(x_n,k\Delta t)\sin(2\pi fk\Delta t), \tag{9}$$

where K_0 is the record length. The cosine and sine coefficients $\{C_n,\ S_n;\ n=0,...,N-1\}$ enable us to reconstruct the complex-valued phase vector P_n at position $x_n=n\Delta$, which is given as

$$P_n=P_{\text{real}}(x_n)+jP_{\text{imag}}(x_n)=C_n-jS_n. \tag{10}$$

Specifically, we chose an equiangle sampling technique, i.e., 512 samples per ten periods (51.2 samples per period). This integer number of samples per ten periods was experimentally found to enable more accurate estimation of the phase vector than other sample sets of noninteger periods. Multiple periods of sampled records were intentionally chosen to reduce both higher- and subharmonic components as much as possible in reconstructing the phase vectors corresponding to the single tone acoustic source. The set of reconstructed multiple phase vectors $\{P_n;\ n=0,1,2,...,6\}$ are used to estimate the best-fitted mean flow velocity according to the procedures described in Eqs. (6)–(8).

V. EXPERIMENTAL RESULTS AND DISCUSSION

To achieve accurate measurement of the phase vectors along the pipe, the calibration procedure of multiple microphones is introduced and calibration results are illustrated. The flow velocity obtained by the proposed acoustic flow meter is compared to that measured by the nozzle-type reference flowmeter. Experimental results carried out to examine the performance of the acoustic flowmeter for different flow conditions are also presented in the following.

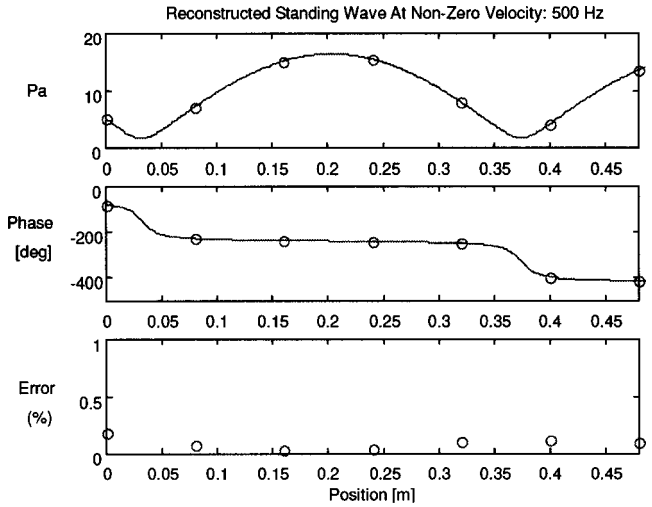


FIG. 5. Spatial acoustic pressure distribution of the standing wave for a 500-Hz pure tone acoustic source.

A. Calibration of microphone array

When multiple microphones, specifically as in this study, are used to reconstruct the standing wave pattern using the phase vectors, finely calibrated responses of the magnitude and phase for each microphone are required.[7,8] The first microphone at the origin ($x=0$) was calibrated at 500 Hz by comparing the primary standard microphone in KRISS. The comparative calibration of other microphones with respect to the first microphone calibrated was carried out using the sound intensity calibrator (B&K 3541+acoustic coupler UA 0914) whose paired ports 1 and 2 provide the equal sound pressure with two microphones. The relative magnitude and phase responses $\{H_{c,n}(f);\ n=1,2,...,6\}$ of the six channels were measured according to ISO 10534-2.[14] The measured magnitude and phase responses were used to compensate the relative magnitude and phase responses using the correction method for paired microphones suggested in ISO 10534-2.

Figure 5 illustrates the reconstructed standing wave pattern for nonzero velocity, whose acoustic pressure levels and phases are compensated by the calibration responses. The symbol ''○'' denotes the measured phase vectors and the solid line does the reconstructed phase vector using the least squared-based standing wave reconstruction method.[15] The percentage error in Fig. 5 was calculated by normalizing the difference between the measured and reconstructed phase vectors by the root-mean-squared acoustic pressure over the one wavelength. It was shown that the maximum error between the measured acoustic pressures and the best-fitted model was about 0.2% for the magnitude and about 0.05° for the phase difference, respectively. This error analysis of acoustic pressure measurements, shown in Fig. 5, was made to check each measurement set of seven phase vectors. Table I illustrates the relative magnitude errors for three different flow control-valve conditions (''closed'' state $U=12.7$ m/s, ''half-open'' state $U=8.6$ m/s, ''fully open'' state $U=4.4$ m/s). It shows that the standard deviation of the relative magnitude errors was less than 0.24% even for the flow control-valve closed, i.e., the fastest velocity condition among them. This measurement accuracy level had been maintained throughout in this work.

TABLE I. Comparison of relative magnitude errors between the measured and reconstructed acoustic phase vectors for three different flow control valve conditions (closed, half-open, fully open).

Microphones	Fully open	Half-open	Closed
Channel 0	0.18%	0.24%	0.23%
Channel 1	0.08%	0.12%	0.16%
Channel 2	0.05%	0.14%	0.09%
Channel 3	0.05%	0.07%	0.09%
Channel 4	0.10%	0.23%	0.23%
Channel 5	0.12%	0.09%	0.14%
Channel 6	0.13%	0.19%	0.22%
Standard deviation	0.13%	0.18%	0.19%

B. Measurement results of mean flow velocity

In this experiment, a 500-Hz pure tone was chosen and its sound level, averaged along the pipe, was 20 dB higher than the acoustic flow noise level. This sound source level was experimentally chosen to obtain the phase vectors with the measurement errors bounded within the range of 0.2%, as noted previously. Furthermore, the steady state of fluid flow in the pipe was also maintained as recommended in ISO/CD 5167,[16] the ratio of the root-mean-squared value of fluctuating pressure component $\Delta P'_{\mathrm{rms}}$ to the time-mean value of the differential pressure ΔP_{mean} was kept at less than 4% ($\Delta P'_{\mathrm{rms}}/\Delta P_{\mathrm{mean}} \approx 0.04$).

Each set of the phase vectors along the pipe, as noted in Sec. IV, was acquired every ten-period interval of the 500 Hz acoustic source such that 50 sets of the measured phase vectors per second were obtained and used to estimate their corresponding flow velocities. This reading rate of flow velocity is very fast in comparison to the reading rate of one to two samples per second typically used in gas flow velocity measurements. To synchronize with the reading rate of the reference flowmeter, 25 consecutive velocity measurements of the acoustic flowmeter for $\frac{1}{2}$ s were averaged to give one sample of its flow velocity. In Fig. 6, the measured flow velocities from the acoustic flowmeter are compared against those measured with the reference flowmeter for 28 different velocity measurement points in the flow velocity range of 2.5–26 m/s. Ten sampled data at each measurement condition were overlapped in Fig. 6. The standard deviation of velocity errors between the measurements of the acoustic flowmeter and the reference flowmeter was observed to be 0.096 m/s. These results may indicate that the linearity of the acoustic flowmeter proposed in this paper is very acceptable to real applications. The linearity is shown to be more accurate than those of previous models by Potzick *et al.*[3] and Alves.[4,5]

The encouraging results presented previously lead to further experimental attempts to examine what amount of measurement accuracy the developed method can present for different environmental conditions; more specifically, for different acoustic pressure patterns along the measurement positions and different flow velocity profiles over the pipe cross-sectional area. Different acoustic pressure distributions

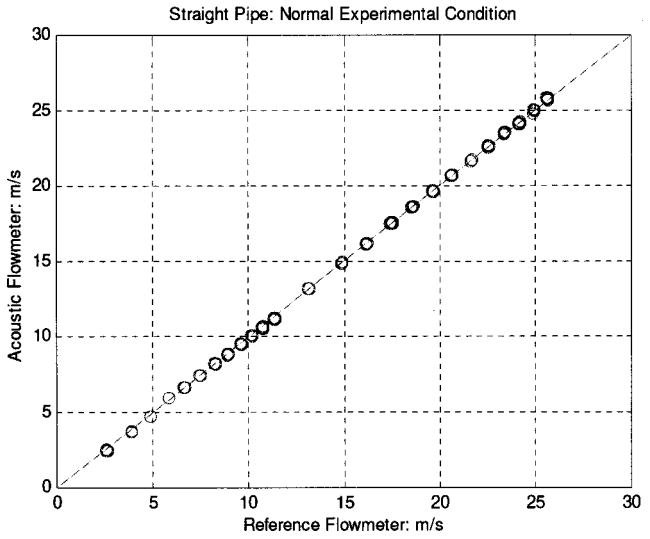


FIG. 6. Comparison of the measured velocities from the acoustic flowmeter (y axis) against those measured from the reference flowmeter (x axis).

along the pipe were made either by connecting two straight pipes of different length (0.4, 0.5 m) to the intake part or removing three straight pipes (0.4, 0.5, 0.5 m) from the intake part. These changes at the intake part make spatial shift of the acoustic pressure pattern along the pipe. Their spatial shift is related to the ratio of the half wavelength to the pipe length added to, or removed from, the intake part. These spatially shifted acoustic pressure patterns at the microphones were exploited to examine their effects on the mean flow velocity measurement. Figure 7(a) shows the experimental results obtained from six different acoustic pressure patterns in the flow measurement section. Ten samples are plotted at each velocity position. The relative measurement errors, i.e., the difference between both flowmeters normalized by the reference one, are less than 2% in the velocity range of $U=7$ m/s. The percentage error in the low velocity range of $U<5$ m/s is larger, i.e., 3%–5%. The measurement results even for different acoustic pressure distributions show a small variation with the standard deviation of 0.112 m/s. Those variations may indicate the measurement resolution of the developed flowmeter. The mean value for different acoustic pressure patterns show a ''curved'' bias from the zero in the velocity range of $U<20$ m/s. This biased feature is regarded as one of the systematic errors for the developed flowmeter. An experimental approach to reduce the velocity bias from zero velocity is in progress.

A fundamental issue in flowmeters[1] has been to achieve an averaged flow velocity over the pipe cross-section area. At the onset of this work, it became apparent that a salient property of the plane wave in the pipe, which is propagated uniformly over the pipe cross-section area with the speed of sound, could be exploited to develop a gas flowmeter expected to be robust in measuring the averaged velocity over the cross-section area. To examine the robustness of the developed acoustic flowmeter, experimental attempts were carried out. In this work, four kinds of pipe elements—a gate valve [denoted by ''elbow-1'' in Fig. 7(b)], a single elbow [denoted by ''elbow-1'' in Fig. 7(b)], closely coupled U-shaped double elbows in plane [denoted by ''elbow-2'' in

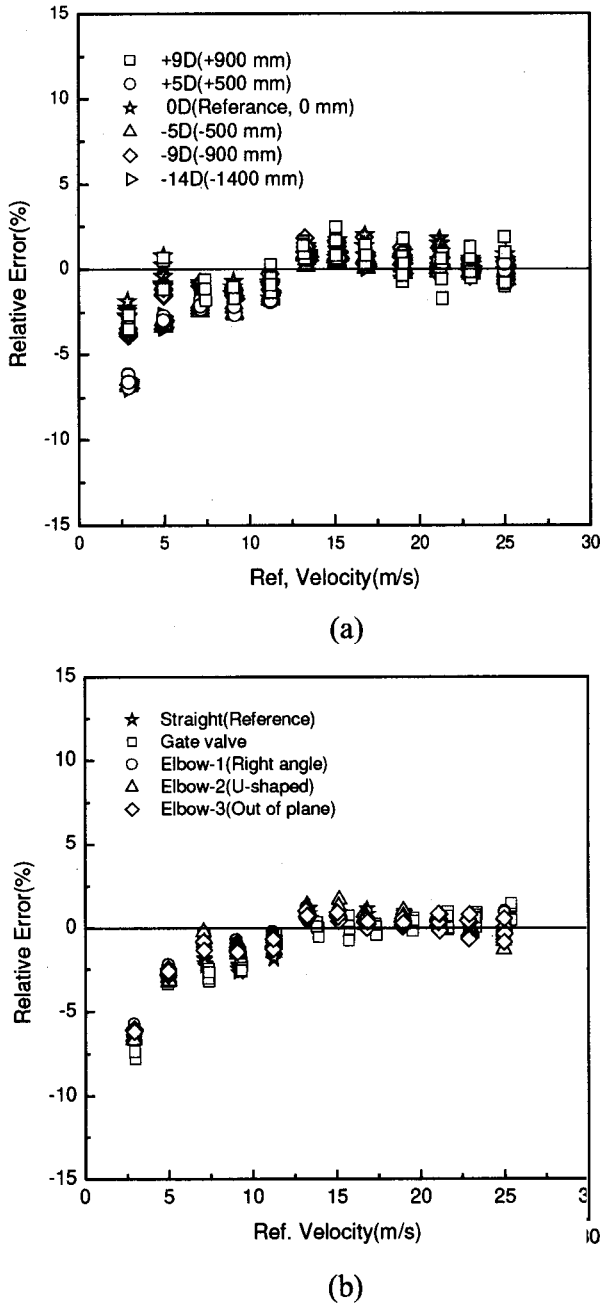


FIG. 7. Relative flow velocity errors between the nozzle-type reference flowmeter and the developed acoustic flowmeter measured from different experimental conditions: (a) for different acoustic pressure patterns (inner diameter $D=100$ mm); (b) for different flow velocity profiles.

Fig. 7(b)], and closely coupled double elbows out of plane [denoted by ''elbow-3'' in Fig. 7(b)]—were chosen to generate different velocity profiles unlike that of the straight pipe. The flow velocity measurement results for five different velocity profiles are shown in Fig. 7(b). The relative measurement errors are seen to be less than 2% in the velocity range of $U\geqslant 7$ m/s regardless of different velocity profiles. Even in the low flow velocity range of $U<5$ m/s, the relative measurement errors decrease in comparison to Fig. 7(a). It seems to be a big advantage for the acoustic flowmeter that exploits the uniformly distributed acoustic plane wave over the pipe across-sectional area to measure the mean flow velocity. The measurements for different velocity profiles are seen to yield

less variation within a finite range than Fig. 7(b), whose standard deviation is equal to 0.101 m/s. It may also be related to the measurement resolution of the developed flowmeter. Still, the mean values for each velocity measurement point reveal a ''curved'' bias from the zero in the range of flow velocity $U<12$ m/s, similar to the results shown in Fig. 7(a). To understand and solve this bias feature, further experimental work is also in progress.

VI. CONCLUDING REMARKS

This paper has presented the theoretical background and experimental verification for a new technique for the measurement of the flow velocity averaged along the finite length of the pipe as well as over the cross section. Experimental data demonstrate the linearity of the proposed acoustic flowmeter in the flow velocity range of up to 27 m/s. Its salient robustness was demonstrated under the different acoustic pressure patterns and different flow velocity profiles over the pipe cross-sectional area. The results have encouraged this research team to further refine the proposed technique to develop a prototype for practical applications. Such refinement will include the minimization of the measurement bias from zero at the low flow velocity and the experimental verification of the proposed method for the different frequencies of the acoustic source.

The use of the multiple equispaced microphones in this work has enabled the measurement of the mean flow velocity in the pipe. The phase vectors of acoustic pressure along the pipe are also shown to provide the incident and reflected acoustic pressures that lead to the direct calculation of reflection coefficient, acoustic impedance, and acoustic absorption coefficients as in ISO 10534. Therefore, the measured flow velocity and the incident and reflected acoustic pressure components in the pipe are expected to provide a way of evaluating the acoustic performances of mufflers under real flow conditions by using only the multiple microphones. Since this method does not distort any acoustic field, it may be more preferable than the nozzle-type gas flowmeter whose installation in the pipe changes the acoustic property itself. Related experimental setups to examine such an advantage are under construction.

ACKNOWLEDGMENTS

This work was partially supported by the Korea Ministry of Science and Technology and the Korea Research Institute of Standards and Science (Project Code Nos. 99-0407-100 and 00-0406-031). The authors would like to acknowledge kind comments and appropriate suggestions made by Professor L.C. Sutherland.

[1] R. W. Miller, *Flow Measurement Engineering Handbook* (McGraw–Hill, New York, 1983).
[2] *Noise and Vibration: P.O.A.L. Davis, Chapter 12. Fundamental Duct Acoustics*, edited by R. G. White and J. G. Walker (Horwood, London, 1982).
[3] J. E. Potzick and B. Robertson, ''Long wavelength acoustic flowmeter,'' U.S. Patent 4,445,389 (1984).
[4] P. S. Alves and A. G. Doige, ''Measurement of mean gas flow velocity in piping using acoustic properties,'' Noise-con 88, West Lafayette, IN, 1988, pp. 469–474.
[5] P. S. Alves, ''Three acoustic flow measurement methods,'' Fourth International Symposium on Fluid Flow Measurement, Denver, CO, 1999.
[6] Y.-D. Chun, Y.-H. Kim, and W.-S. Cheung, ''The measurements of acoustic field and mean flow velocity in pipes using microphone array,'' J. KSME **22**, 1761–1768 (1998).
[7] W.-S. Cheung and J.-S. Paik, ''A new acoustic flowmeter measuring the mean pipe flow velocity,'' in Ref. 5.
[8] W.-S. Cheung *et al.*, ''Development of the mean fluid velocity measurement technique using the sound field reconstruction,'' KRISS Report No. KRISS-IR-2000-015, 2000.
[9] ISO 10534-1, *Acoustics—Determination of Sound Absorption Coefficient and Impedance in Impedance Tubes. 1. Method Using Standing Wave Ratio* (International Organization for Standardization, Geneva, Switzerland, 1996).
[10] L. L. Beranek, *Acoustics*, 2nd ed. (The Acoustical Society of America, 1986).
[11] G. Strang, *Linear Algebra and its Applications*, 3rd ed. (Harcourt Brace Jovanovich, 1988).
[12] The Math Works, Inc, MATLAB Reference Guide (Ver. 5.3), 1999.
[13] M. R. Spiegel, *Mathematical Handbook of Formulas and Tables*, Schaum's Outline Series (McGraw–Hill, New York, 1968).
[14] ISO 10534-2, *Acoustics—Determination of Sound Absorption Coefficient and Impedance in Impedance Tube. 2. Transfer Function Method* (International Organization for Standardization, Geneva, 1998).
[15] W.-S. Cheung, M.-J. Jho, and Y.-H. Kim, ''Improved method for the measurement of acoustic properties of a sound absorbent sample in the standing wave tube,'' J. Acoust. Soc. Am. **97**, 2733–2739 (1995).
[16] ISO/CD 5167-1, *Measurement of Fluid Flow by Means of Pressure Differential Devices Inserted in Circular Cross-section Conduits Running Full. 1. General* (International Organization for Standardization, Geneva, Switzerland, 1999).

Some aspects of sound extinction by zooplankton

N. Gorska[a)]
Institute of Oceanology, Polish Academy of Sciences, Powstańców Warszawy 55, 81-712 Sopot, Poland

D. Chu[b)]
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 1 June 2000; revised 30 April 2001; accepted 12 July 2001)

To address the importance of sound extinction (or the shadowing effect) in the presence of a densely aggregated zooplankton layer, acoustic scattering by weakly scattering fluid objects is studied theoretically. An analytical expression for the extinction cross section is obtained based on the forward scattering theorem and the analytical formula for the forward scattering amplitude is derived from the Modal Based–Deformed Cylinder Model (MB–DCM). The validity of the MB–DCM solution for the forward scattering amplitude and its sensitivity to geometrical and physical parameters is studied. Comparison with the PC–DWBA (Phase-Compensated–Distorted Wave Born Approximation) shows a reasonable agreement between the two models. The extinction cross section is shown to be proportional to $(ka)^2$ and the induced attenuation is predicted for various species of zooplankton over a wide frequency range. It is shown that, under certain realistic conditions, such as the presence of a swarm of aggregated krill, sound extinction by zooplankton could influence the acoustic measurements significantly, as much as an 85% reduction in acoustic intensity for *Euphausia superba* with $n=5000$ ind./m^3 over a range of 50 m. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1400738]

PACS numbers: 43.20.Jr, 43.30.Ft [DLB]

I. INTRODUCTION

Acoustic techniques enable us to sample oceans much more efficiently than conventional techniques, such as trawls and nets, and can be used to remotely estimate the biomass of biota living in the ocean. In general, scattering-induced attenuation does not have significant impact on the received acoustic scattering signals and is ignored under normal conditions. However, when biological scatterers are densely aggregated or distributed on an extended scale, a knowledge of sound attenuation by biological objects is necessary to provide an accurate interpretation of the acoustic measurement.[1]

Attenuation of sound energy caused by fish aggregation is important in fisheries acoustics and has attracted much attention.[2–14] Such attenuation is related to, and can be expressed in terms of the acoustic extinction cross section.[1] Various methods of measuring fish extinction cross sections have been developed, including the methods of direct measurement,[2–7] indirect measurement,[8,9] and the method of reference targets.[1,10–13] With these methods, extinction measurements have been made on various fish species and over a wide frequency range.[1–13] Theoretical analyses have been sought to characterize these measurements and analytical expressions of the fish extinction cross section have been derived.[3,14]

In contrast to fish scattering, scattering-induced attenuation has not attracted much attention in zooplankton acoustics since zooplankton are, in general, weaker scatterers and their sound extinction is usually negligible. However, it has been theoretically demonstrated that the extinction in the presence of densely aggregated zooplankton can be noticeable.[15,16] A number of publications report that, based either on biological sampling or photo picture analysis, the density of a swarm of euphausiids varies from hundreds ind./m^3 to tens of thousands ind./m^3,[17,18] and can be as high as 100 000 ind./m^3.[19] In these cases, sound extinction is expected to be quite important in interpreting the acoustic data. Sound extinction is also important in measuring zooplankton sound speed[20–22] since the volume concentration of zooplankton can be as high as 50% in order to obtain a more reliable measurement.[22]

Gruber and Meister conducted laboratory measurements of sound extinction by brine shrimp at megahertz frequencies.[23] They found that the attenuation coefficient varies as $f^{1.35}$, where f is the acoustic frequency. To our knowledge, field data have not been published studying sound attenuation by zooplankton.

The forward scattering theorem has been used to describe the sound extinction by biological targets.[14–16] According to the forward scattering theorem,[24] the extinction cross section of an individual animal, σ_{ext}, which describes the sound attenuation,[25] can be expressed as

$$\sigma_{\text{ext}}=(4\pi/k)\,\text{Im}\,f(\hat{r}_i,\hat{r}_i), \tag{1}$$

where $f(\hat{r}_i,\hat{r}_i)$ refers to the forward scattering amplitude, Im means the imaginary part, $k=2\pi\nu/c$ denotes the wave number, where c is the sound speed in seawater, and ν is the acoustic frequency. The vector $\hat{r}_i$ is the unit position vector of the incident wave. To use Eq. (1), scattering models that describe $f(\hat{r}_i,\hat{r}_i)$ are required. Both analytical and numerical methods have been developed to study bistatic scattering.[14–16,26–29] The former provides the exact and approximate solutions for targets of simple shapes, while the latter can be used for targets of more coplicated shapes. Reviews of these methods can be found in Refs. 26–29. Two of

[a)]Electronic mail: gorska@iopan.gda.pl
[b)]Electronic mail: dchu@whoi.edu

these methods, the Phase-Compensated–Distorted Wave Born Approximation (PC–DWBA) and the Modal Based–Deformed Cylinder Model (MB–DCM), have been employed to the extinction problem in previous studies of extinction.[14–16]

The MB–DCM is based on the exact modal series solution for scattering by an infinitely long cylinder[30] and has been successfully used to analyze backscattering by zooplankton of various species.[30–33] The model is applicable to targets with a variety of material properties (fluid, solid elastic, elastic shell, etc.), and is quite accurate as long as the aspect ratio of the cylinder is reasonably high and the geometrical and physical properties along the cylinder axis vary slowly. The MB–DCM is restricted to the case where the angles of incidence and scattering are normal or near normal (broadside). The range of off-broadside orientations, within which the MB–DCM is still accurate, has been defined based on a careful study of the model over a wide range of shapes and material properties of scattering targets.[27,29]

The PC–DWBA model is applicable to sound scattering by fluid-like, weakly scattering targets.[15] In contrast to the MB–DCM, the PC–DWBA can be used for arbitrary shapes and orientations of scatterers and can provide simpler analytical expressions for both backward and forward scattering problems.

In a recent paper,[16] the MB–DCM was employed to evaluate the sound extinction by several krill species. Bent cylinders were used to describe the geometrical shape of individual krill. The application of the PC–DWBA to zooplankton scattering has been demonstrated in Ref. 15. The analytical solution for the forward scattering amplitude based on the PC–DWBA has been obtained and its sensitivity has been studied.

In this paper, a systematic analysis is presented to provide more information on the sound extinction by zooplankton. In Sec. II, analytical expressions for the forward scattering amplitude of individual objects are derived on the basis of the MB–DCM. Various geometrical shapes frequently used in modeling sound scattering by elongated individual zooplankton are considered, including straight cylinders, bent cylinders, and prolate spheroids. The accuracy of the MB–DCM solution in the forward scattering case is studied and its sensitivity to the geometrical and acoustical parameters of individual zooplankton is analyzed. Comparison between the forward scattering amplitude based on the MB–DCM and that based on the PC–DWBA is provided. Similar analysis for backscattering is also presented to understand the fundamental difference of the acoustic scattering characteristics between forward scattering and backscattering. In Sec. III, the analytical expressions for the extinction cross section averaged over geometrical and acoustical parameters of zooplankton are presented to describe sound attenuation by densely aggregated zooplankton. Both the MB–DCM and the PC–DWBA approaches are employed. The significance of extinction for densely aggregated zooplankton is demonstrated. Finally, conclusions are drawn in Sec. IV.

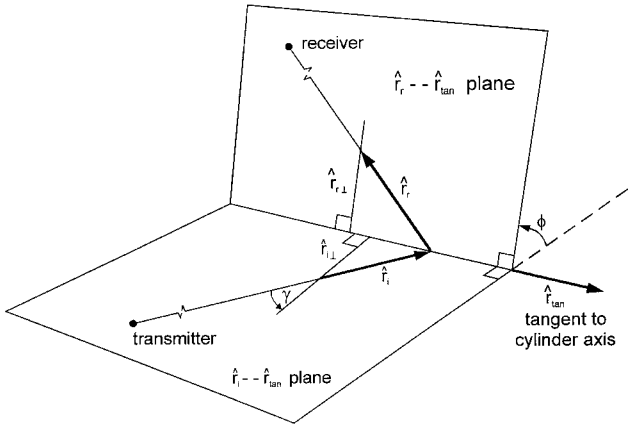


FIG. 1. Geometry of the bistatic scattering by a deformed cylinder.

II. EXTINCTION CROSS SECTION OF INDIVIDUAL ZOOPLANKTON

A. Analytical expressions for forward scattering

1. Modal based deformed cylinder model (MB–DCM)

The scattering amplitude $f(\hat{r}_i,\hat{r}_r)$ of a finite deformed cylinder can be written as[30]

$$f(\hat{r}_i,\hat{r}_r)=\int_{r_{\rm pos}}\sum_{m=0}^{\infty} b_m\cos(m\Phi)\times\exp[i\kappa r_{\rm pos}(\hat{r}_i-\hat{r}_r)\cdot\hat{r}_{\rm pos}]|d\bar{r}_{\rm pos}|, \tag{2}$$

where the scattering coefficient b_m is the coefficient derived from the exact modal series solution for an infinitely long cylinder. The scattering geometry is illustrated in Fig. 1 taken directly from Ref. 30. Here $\hat{r}_i$ and $\hat{r}_r$ are unit vectors describing the directions of the incident and received (scattered) waves, $\bar{r}_{\rm pos}=r_{\rm pos}\hat{r}_{\rm pos}$ is the position vector of an arbitrary point on the axis of the cylinder, where $\hat{r}_{\rm pos}$ denotes a unit position vector originated from the origin of the coordinate system. The azimuthal angle Φ is the angle between $\hat{r}_{i\perp}$ and $\hat{r}_{r\perp}$ and $\Phi=\cos^{-1}(\hat{r}_{i\perp},\hat{r}_{r\perp})$ ($\Phi=0$ for forward scattering and $\Phi=\pi$ for backscattering). The unit vector $\hat{r}_{i\perp}$ is in the $\hat{r}_i-\hat{r}_{\rm tan}$ plane and is perpendicular to $\hat{r}_{\rm tan}$. Similarly, $\hat{r}_{r\perp}$ is in the $\hat{r}_r-\hat{r}_{\rm tan}$ plane and is also perpendicular to $\hat{r}_{\rm tan}$. $\hat{r}_{\rm tan}$ is the unit vector tangent to the cylinder axis.

The coefficient b_m can be written in the following form:[30]

$$b_m=-\frac{\varepsilon_m}{1+iC_m}, \tag{3}$$

where the Neumann factor $\varepsilon_m=1$ for $m=0$, and $\varepsilon_m=2$ for $m>0$, and the coefficient C_m can be expressed as

$$C_m=\frac{[J_m'(k'a)N_m(ka)]/[J_m(k'a)J_m'(ka)]-gh[N_m'(ka)/J_m'(ka)]}{[J_m'(k'a)J_m(ka)]/[J_m(k'a)J_m'(ka)]-gh}, \tag{4}$$

where a denotes the cylindrical radius of the deformed cylinder at $\vec{r}_{\text{pos}}$, $k=\kappa\cos\gamma$ with γ being the angle between the directions of the incident wave and the axis normal of the cylinder at $\vec{r}_{\text{pos}}$ (Fig. 1). $k=k/h$ is the wave number inside the scatterer. $J_m(x)$ and $N_m(x)$ are the Bessel functions of the first and second (Neumann) kinds of order m, respectively, and $J'_m(x)$ and $N'_m(x)$ are the corresponding first order derivatives with respect to x. $h=c'/c$ and $g=\rho'/\rho$ are the sound speed and density contrasts, respectively. The superscript "$'$" in Eq. (4) refers to the quantities inside the scattering object, ρ and c are the density of and sound speed in the surrounding water.

It should be pointed out that this solution is more appropriate for normal or near normal incidence, and for an aspect ratio much larger than unity. In addition, the solution requires slow variations in the radius of curvature, the cross sectional area, and the material properties (g and h) along the length-wise axis of the deformed cylinder.

For forward scattering, $\hat{r}_i=\hat{r}_r$ and $\Phi=0$, Eq. (2) can be simplified to

$$f(\hat{r}_i;\hat{r}_i)=\int_{r_{\text{pos}}}\sum_{m=0}^{\infty} b_m|d\hat{r}_{\text{pos}}|. \tag{5}$$

This expression is a general expression for describing forward scattering by an axially symmetric elongated object. Solutions with various geometrical shapes, such as a straight cylinder, a bent cylinder, and a prolate spheroid considered in Ref. 30 for the backscattering case, can be easily obtained.

Straight cylinder: For a straight cylinder of length l [Fig. 2(a)], the integration over $|\vec{r}_{\text{pos}}|$ for broadside incidence is replaced by the integration over length along the cylinder axis. In the case of a constant composition profile along the axis, the modal series can be taken out of the integral and the forward scattering amplitude can be expressed as

$$f(\hat{r}_i,\hat{r}_i)=-\frac{il}{\pi}\sum_{m=0}^{\infty} b_m, \tag{6}$$

where the scattering angle γ [Fig. 2(b)] equals 0 for broadside incidence.

Uniformly bent cylinder: For a homogeneous (constant material properties, g and h), uniformly bent (constant radius of curvature, ρ_c), and untapered cylinder (constant cross-sectional radius a), the forward scattering amplitude $f(\hat{r}_i,\hat{r}_i)$ can be simplified by inserting $|d\vec{r}_{\text{pos}}|=\rho_c d\gamma$ and replacing the integration in Eq. (5) with an integration over the angle γ [Fig. 2(b)]:

$$f(\hat{r}_i,\hat{r}_i)=-\frac{i}{\pi}\frac{l}{\gamma_{\max}}\int_0^{\gamma_{\max}} d\gamma\sum_{m=0}^{\infty} b_m, \tag{7}$$

where $\gamma_{\max}=l/2\rho_c$ with l is the total arc length of the bent cylinder.

Prolate spheroid: To apply Eq. (5) to a prolate spheroid, we introduce a tapering function, $a(z)$, defining the change of the radius along the cylindrical axis,

$$a(z)=a[1-(z/(l/2))^2]^{1/2},$$

where a is the semiminor axis, $l/2$ is the semimajor axis, and z is the coordinate along the major axis centered at the center of the spheroid [Fig. 2(c)]. Substituting $u=z/(l/2)$ and $du=dz/(l/2)$ in Eq. (5), and assuming broadside incidence (normal to the major axis), the forward scattering amplitude can be expressed as

$$f(\hat{r}_i,\hat{r}_i)=-\frac{i}{\pi}l\int_0^1 du\sum_{m=0}^{\infty} b_m. \tag{8}$$

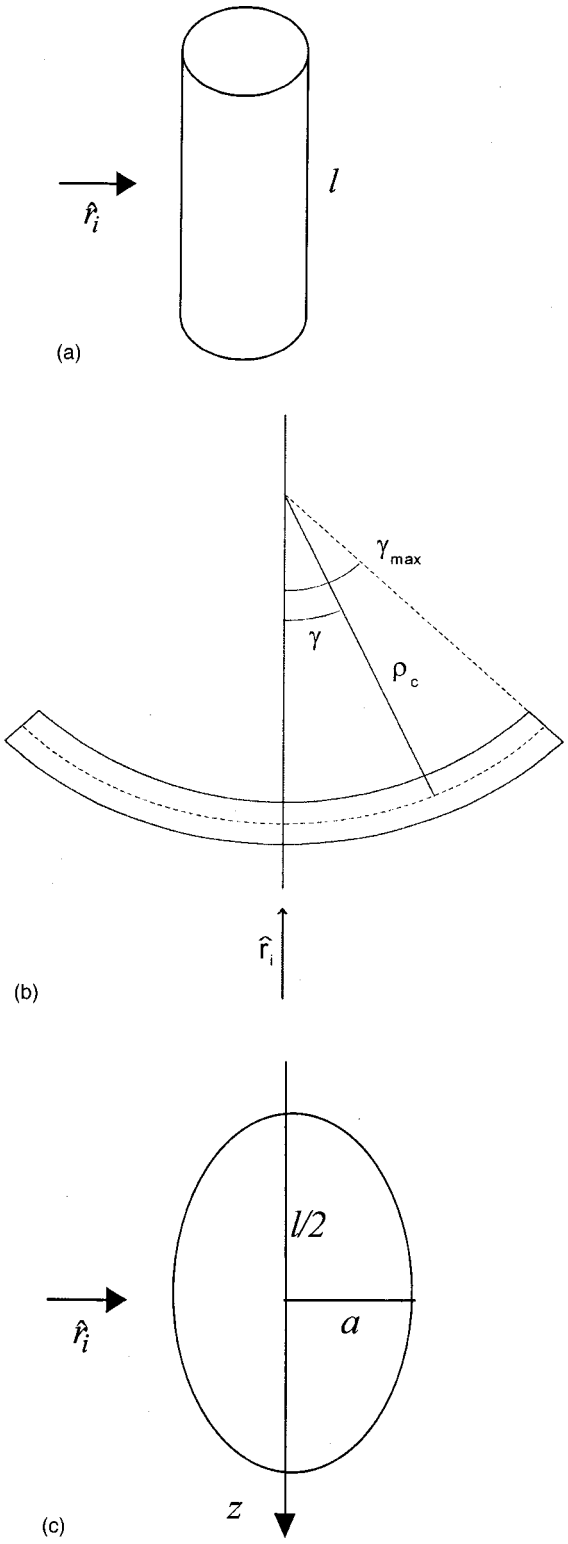


FIG. 2. Geometry of the forward scattering by a straight cylinder (a), a uniformly bent cylinder (b), and a prolate spheroid (c).

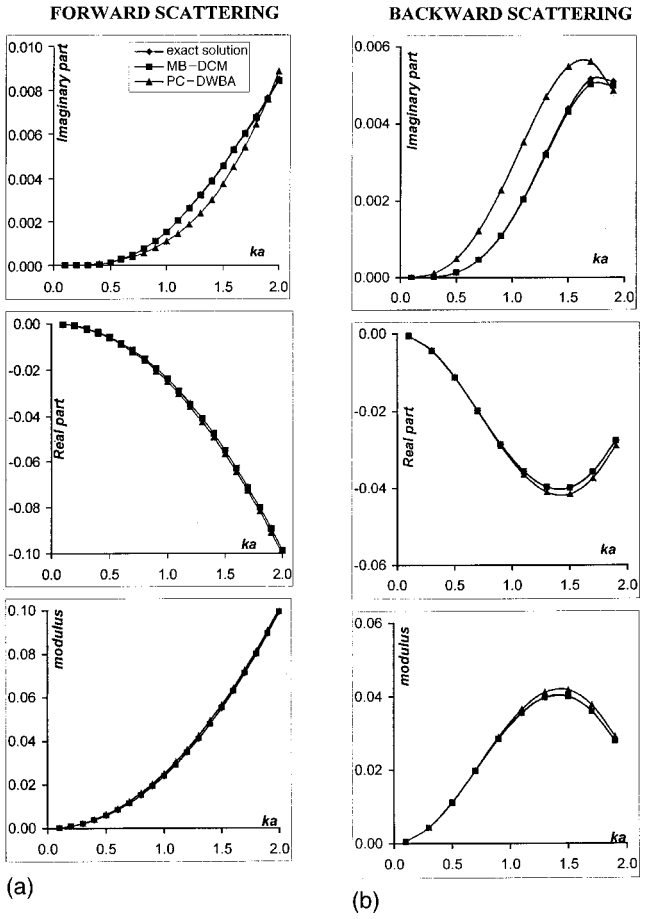


FIG. 3. Comparison of the scattering amplitude of a prolate spheroid normalized by the semimajor axis for the MB–DCM, the PC–DWBA and the exact modal series solutions. The results for the forward scattering and the backscattering are presented in the left (a) and right columns (b), respectively. The curves marked by diamonds, squares, and triangles correspond to the exact, MB–DCM, and PC–DWBA solutions, respectively. For the exact and the MB–DCM solutions, the discrepancies are indistinguishable. The simulation parameters are $g=h=1.04$, and $e=5$.

The radius a implicitly included in b_m [Eqs. (3) and (4)] can be expressed in terms of a dimensionless variable u as $a(u)=a[1-u^2]^{1/2}$ and the angle γ is set to 0 for broadside (normal) incidence.

To evaluate the accuracy of the model, we compare the MB–DCM based solution (diamond) with the exact modal series solution (square)[34] for the forward scattering [Eq. (8)] and backscattering amplitudes[30] of a prolate spheroid (Fig. 3). Previous accuracy analysis of the MB–DCM focused on the modulus of the scattering amplitude and did not address the problem of extinction.[27,29,30] To study the extinction problem, validations of the imaginary part of the forward scattering amplitude obtainable from the MB–DCM based solutions are necessary. A comparison between the exact and MB–DCM based solutions for forward scattering [Fig. 3(a)] and backscattering [Fig. 3(b)] is presented in Fig. 3. In the figure, the imaginary part (top row), the real part (mid row), and the modulus (bottom row) of the scattering amplitude are normalized by the semimajor axis. The parameter ka of the horizontal axis is a dimensionless parameter, where k is the wave number and a is the semiminor axis. The simulation parameters used in obtaining Fig. 3 are $g=h=1.04$, the aspect ratio $e=5$ (ratio of the semimajor axis to the semiminor axis), and the incidence is broadside.

The agreement between the MB–DCM and the exact solution are excellent. The two curves coincide almost exactly and the differences between the two curves are hardly visible.

2. Phase-compensated–distorted wave born approximation (PC–DWBA)

The PC–DWBA is based on the conventional DWBA approach,[32,33] with an additional multiplicative factor to approximately account for the phase mismatch between the exact solution and the DWBA based solution. This term is a function of shape and the dimensionless parameter ka. The PC–DWBA representation of the scattering amplitude of a fluid-like prolate spheroid has been obtained in Ref. 15, Eq. (22):

$$f=(k')^2a^3e(h^2\gamma_k+\gamma_\rho\cos\theta_s)\frac{j_1(k'a\Phi(\theta_s,e))}{k'a\Phi(\theta_s,e)}\times\exp(-i\tfrac{2}{3}[1+\sin(\theta_s/2)](h-1)k'ae^{1/3}), \tag{9}$$

where θ_s denotes the angle between the incident and scattering directions [see Fig. 8(a) from Ref. 15] and $j_1(x)$ is the spherical Bessel function of order 1. $e=l/2a$ describes the aspect ratio, where l is the length of the prolate spheroid. Other variables in Eq. (9) are

$$\Phi(\theta,e)=[(1-\cos\theta_s)^2+e^2\sin^2\theta_s]^{1/2},$$

$$\gamma_k=(1-gh^2)/(gh^2),\quad \gamma_\rho=(g-1)/g.$$

To obtain simpler analytical expressions for forward scattering and backscattering, we can express the spherical Bessel function in terms of the elementary functions[35] and the coefficients γ_k and γ_ρ in terms of g and h. Equation (9) can then be expressed as

$$f=\frac{(k')^2V}{4\pi}(h^2-1)\exp[-i\alpha_s(h-1)k'a]\approx\frac{(k')^2V}{4\pi}(h^2-1)[1-i\alpha_s(h-1)k'a], \tag{10}$$

for the forward scattering ($\theta_s=0$) and

$$f_{\rm bsc}=\frac{l}{8}\left[\frac{2}{g}-h^2-1\right]\left[\frac{\sin(2k'a)}{2k'a}-\cos(2k'a)\right]\times\exp[-2i\alpha_s(h-1)k'a] \tag{11}$$

for backscattering ($\theta_s=\pi$). $\alpha_s=(2/3)e^{1/3}$ is the shape parameter and V represents the volume of the scatterer.

In Eq. (10), the second part presented in the second line is a simplification of the first part for $\alpha_s[h-1]k'a\ll 1$. For fluid-like scatterers, errors resulting from such an approximation are sufficiently small for small ka since h is close to unity. They are only 0.14% at $ka=2$ and 3.5% at $ka=10$ for the imaginary part of forward scattering amplitude. For the real part, the errors are 0.42% at $ka=2$ and 10.4% at $ka=10$, respectively.

The results computed from the PC–DWBA model (triangle) are superimposed in Fig. 3. The comparison shows a

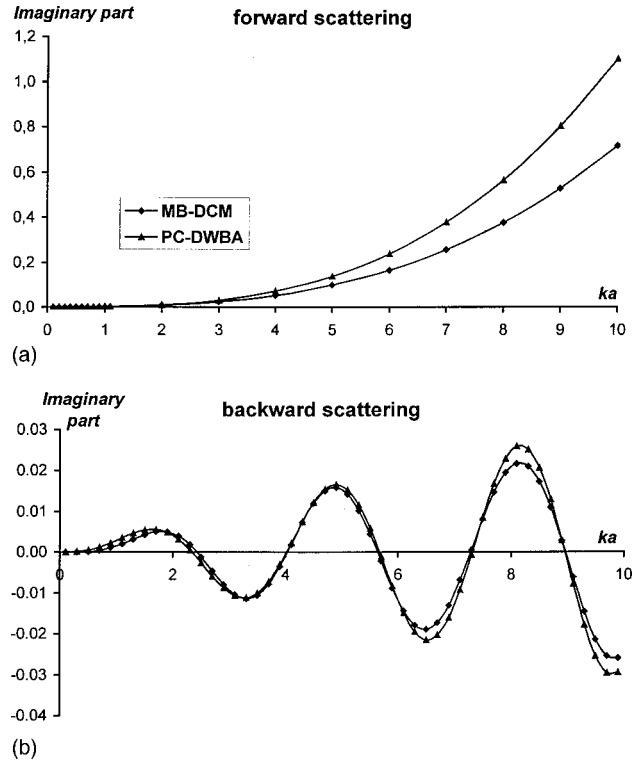


FIG. 4. Comparison of the forward scattering (a) and the backscattering (b) amplitudes between the MB–DCM and the PC–DWBA solutions for a prolate spheroid. The lines marked by diamonds are obtained from the MB–DCM expressions: from Eqs. (3), (4), and (8) for the forward scattering, and from the expressions given in Ref. 30 for the backscattering. The curves marked by triangles are computed from the PC–DWBA solution: Eq. (10) for the forward scattering and Eq. (11) for the backscattering. The simulation parameters are $g=h=1.04$, and $e=5$.

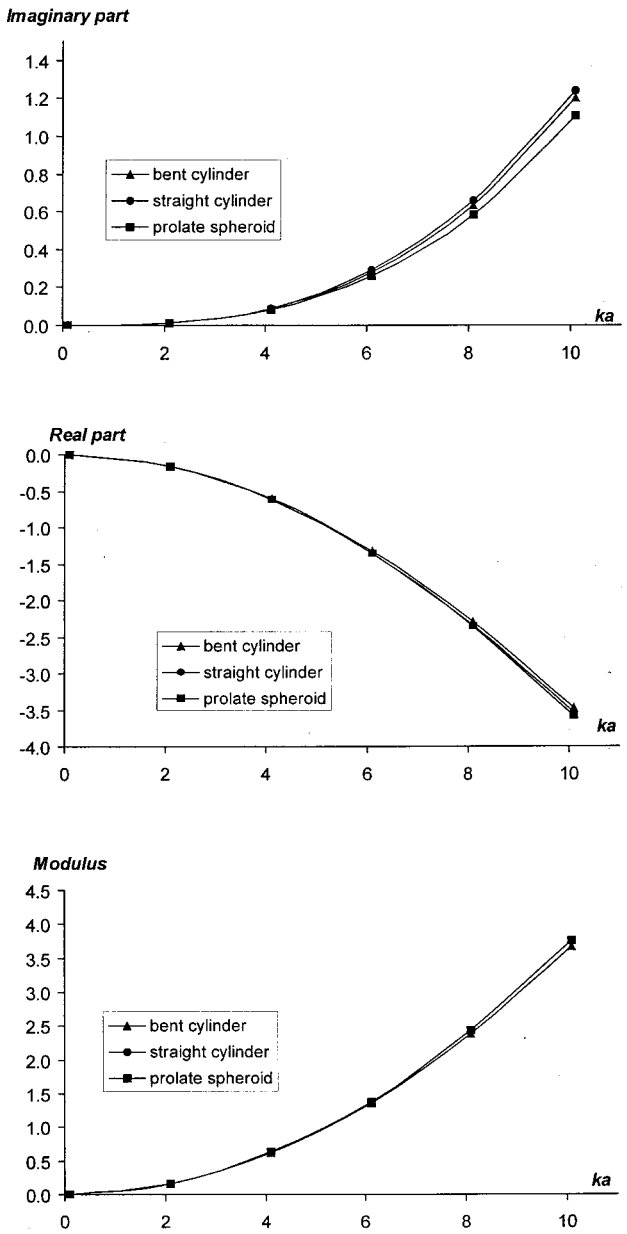


FIG. 5. Influence of the geometrical shape of individual scatterers on the normalized forward scattering amplitude. Computations corresponding to bent (triangles) and straight (circles) cylinders, and prolate spheroid (squares) are based on the MB–DCM solutions, Eqs. (6), (7), and (8), respectively. The simulation parameters are $g=h=1.04$ for all three shapes and $\rho_c/l=2$ for the bent cylinder.

reasonable fit between the MB–DCM and the PC–DWBA solutions for the real parts. The small difference is indicated only at $1<ka<2$ for the backscattering case.

It can be seen that the imaginary parts of the scattering amplitudes, which are an order of magnitude smaller than the corresponding real parts, are more sensitive to the choice of models than the real parts. For broadside incidence, the MB–DCM fits the exact modal series solution better than the PC–DWBA result and is expected to provide better estimates of the extinction cross section. However, the MB–DCM requires intensive computation time since it uses more complicated analytical expressions, Eqs. (5)–(8), with the coefficients given by Eqs. (3) and (4).

The comparison between the two models for the more sensitive imaginary part at extended *ka* range is shown in Fig. 4. The other simulation parameters are kept unchanged as for Fig. 3. It is observed that the difference grows with increasing *ka* for both the forward and the backscattering amplitudes. Both models predict approximately $(ka)^3$ dependence, or a $(ka)^2$ dependence for the normalized extinction cross section [Eq. (1)].

Although the model is not perfect, the simple PC–DWBA based solution enables us to obtain some physical insights into the scattering problem. First of all, it can be deduced that the forward scattering amplitude is insensitive to the shape and orientation of the scatterer. Second, the forward scattering amplitude is more sensitive to the sound speed contrast than to the density contrast. Third, the forward scattering amplitude is proportional to the volume of the scatterer and to the frequency (wave number) squared.

B. Sensitivity analysis

This section investigates the dependence of the forward scattering amplitude on various parameters by presenting a systematic sensitivity analysis, which will help to further understand the fundamentals of sound extinction phenomenon.

To study shape dependence, we compare the forward scattering amplitudes of three different targets (Fig. 5): the straight finite cylinder, Eq. (6), the uniformly bent cylinder, Eq. (7), and the prolate spheroid, Eq. (8). To make a mean-

ingful comparison, we maintain the volume of the scattering objects to be the same. In order to have an equal volume, the forward scattering amplitude is normalized by $l/2$ for both bent and straight cylinders and by $(3/2)*(l/2)=3/4l$ for spheroids. The imaginary part (top plot), the real part (middle plot), and the modulus (bottom plot) of the normalized forward scattering amplitudes are presented on the vertical axis. The three lines in each plot correspond to the bent cylinder, the straight cylinder, and the prolate spheroid. The calculation parameters are $g=h=1.04$, $l/a=10$, and $\rho_c/l=2$ (bent cylinder). The results demonstrate that the relative differences of the normalized amplitudes are not more than 2.5% for the moduli and the real parts, and are less than 20% for the imaginary part. This observation indicates a slight dependence of the forward scattering amplitude on the geometrical shape of the scatterers and is consistent with predictions from Eq. (10).

According to Ref. 30, the situation is different for the backscattering case and the discrepancy of the normalized backscattering cross sections for sphere, straight, and bent cylinders is observed in the geometric scattering region $ka \geqslant 1$.[30] Following the expressions given in Ref. 30, the backscattering amplitudes for a straight cylinder, a bent cylinder, and a prolate spheroid of equal volume can be calculated. The results are presented in Fig. 6 for the imaginary part (top plot), the real part (middle plot), and the modulus (bottom plot) of the backscattering amplitude. In obtaining Fig. 6, we have used the same parameters as in Fig. 5. For backscattering by a prolate spheroid and a straight cylinder, the normalized scattering amplitude for broadside incidence is independent of the aspect ratio. Again, the amplitudes are normalized as in the forward scattering case described in the previous paragraph. There are many interesting features observable from the figure. For the imaginary part, the difference among the curves is most significant over the entire range of *ka*. For the real part and the modulus the discrepancies among the various scattering objects are indistinguishable on the current scale in the Rayleigh scattering region $ka \ll 1$. However, for $ka > 1$, the overall difference is much larger than in the forward scattering case (Fig. 5) since the backscattering is a geometrical shape-related scattering process.[36]

The fact that the backscattering amplitude is much more sensitive to the shape of the scatterer than the forward scattering can be explained with the following arguments.[31,32] For backscattering, the phase difference between the scattered wave from the front interface and the back interface is $2k'a$, where k' is the wave number inside the scatterer and a is the dimension of the scatterer in the direction of the incidence. Based on the Huygens principle, wavelets from every portion of the scatterer contribute to the total scattered field and add up coherently. If $2k'a$ is equal to 2π or its multiples, the waves from the front and back interfaces are in phase and tend to double the amplitude (weakly scattering case). However, if $2k'a$ is equal to π or its multiples, the waves from the front and back interfaces are out of phase and tend to cancel each other out. If $2k'a$ falls in between, the waves are summed up either constructively or destructively, depending on the size, shape, and orientation of the target, as well as on the wave number inside the scatter (k'). However, for the forward scattering, the phase difference between the waves propagating through and not through the scatterer is $2\Delta ka$, where Δk is the difference between the wave numbers in the surrounding water and in the scatterer. For weakly scattering objects, as long as *ka* itself is not very large, $\Delta ka=\Delta hka$ is very small, and hence wavelets from every part of the scatterer tend to add up constructively regardless of the shape and orientation of the scatterer.

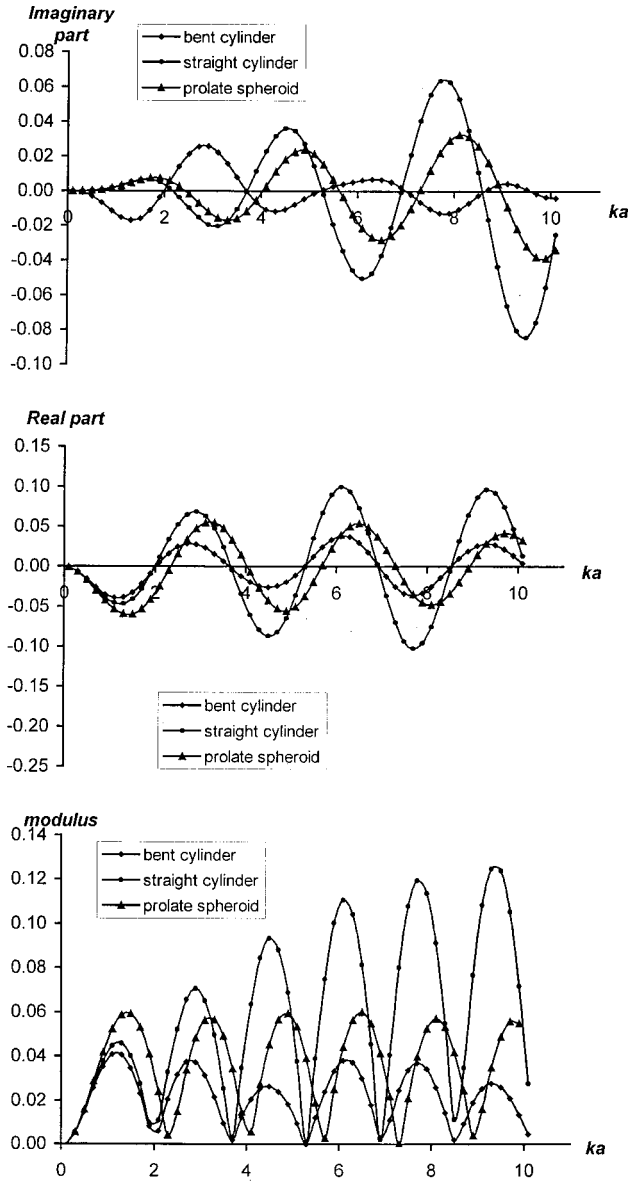


FIG. 6. Influence of the geometrical shape of the scatterer on the normalized backscattering amplitude. Computations corresponding to bent (triangles) and straight (circles) cylinders, and prolate spheroid (squares) are based on the MB–DCM solutions, Eqs. (6), (7), and (8), respectively. Simulation parameters are $g=h=1.04$, $l/a=10$ for the straight and bent cylinders and $e=5$ for the prolate spheroid. For the bent cylinder the ratio is $\rho_c/l=2$.

For bent cylinders, the influence of the radius of curvature of the longitudinal axis on the forward scattering amplitude is presented in Fig. 7, where similar contrast parameters are used as described for Fig. 5. The four numbers given in the legends in each plot correspond to various ratios of the radius of curvature to the length of the cylinder. For all three plots, the curves are so close that they are almost indistin-

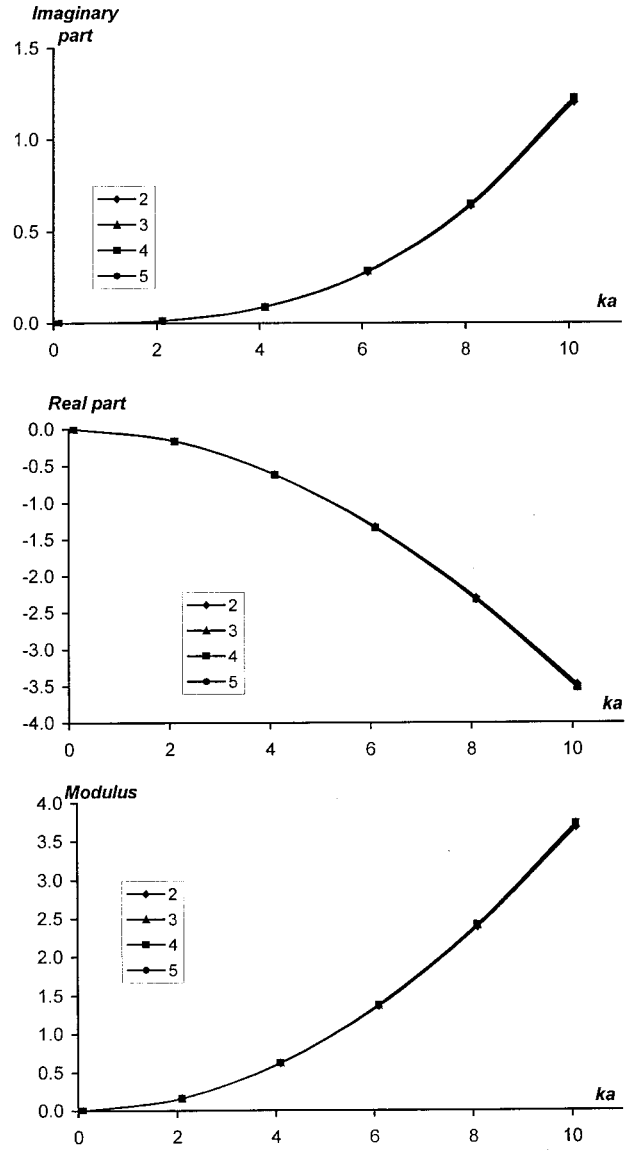


FIG. 7. Sensitivity of the normalized forward scattering amplitude to the parameter ρ_c/l. Computations are based on the MB–DCM solution, Eq. (7). The values of ρ_c/l are indicated in the legend. Four curves are so close that they are almost indistinguishable. The simulation parameters are $g=h=1.04$.

guishable, suggesting that the forward scattering is insensitive to the target shape. In contrast, the backscattering amplitude does moderately depend on the radius of curvature[30] and shows a different characteristic.

In the previous analysis,[15,16] we have not considered the influence of sound absorption inside the scatterer on the forward scattering amplitude. This is due to a lack of readily available data and insignificant sound absorption in zooplankton. However, to perform a systematic sensitivity analysis, it is logical to study the influence of tissue absorption on the acoustic scattering.

To evaluate the sound absorption, following Vogt *et al.*,[37] we replace the real wave numbers k' inside the animal body in the coefficients C_m [Eq. (4)] by the complex wave numbers, $k_a'=k'(1+i\varepsilon)$. Here ε is the absorption coefficient. For the current analysis, we choose ε to vary from 10^{-8} to 10^{-4}. Such a choice is based on the absorption coefficient in seawater over an acoustical frequency range 30–300 kHz.[38] This is reasonable since the acoustic properties of zooplankton tissue and the ambient water are very similar (g and h are close to 1). Inserting k_a' into Eq. (8) and using the same contrast parameters $g=h=1.04$, the imaginary part of the forward scattering amplitude as a function of ε at two ka values, for $ka=5$ and $ka=0.5$ can be computed. The results (not shown) indicate that the imaginary part is not sensitive to the ε over this range and is equal to 1.5×10^{-4} and 0.098 for $ka=0.5$ and $ka=5$, respectively. This result is consistent with what has been found for backscattering[31] and what we expected. For backscattering, even incorporating absorption, the strong interference pattern of the backscattering target strength can still be observed, suggesting that the echoes from the front and back interfaces of the animal have a similar scattering amplitude, hence implying weak absorption in the tissue. In addition, the good agreement between the measured data and the models that neglect absorption in the tissue also indicates an insignificant absorption.[31]

We will not present the sensitivity analysis of the forward scattering amplitude to g and h contrast parameters since these dependencies have already been studied in detail.[15] It was found that the real part of the forward scattering amplitude is approximately proportional to the value $\Delta h=h-1$, while the dependence of the imaginary part on Δh is approximately quadratic.[15] Furthermore, the forward scattering amplitude is weakly dependent on the density contrast parameter g.

C. Examples

In this section, two examples are presented to illustrate the effects of scattering-induced attenuation. To compute the extinction cross section, we apply the forward scattering theorem, Eq. (1), and choose a prolate spheroid as a representative shape for individual zooplankton. The expression for the forward scattering amplitude of a prolate spheroid given by Eq. (8) is employed and the contrast parameters $g=h=1.04$ and aspect ratio $e=5$ are used. The dependence of the quantity $10\log_{10}(\sigma_e/l^2)$ on the parameter ka is plotted in Fig. 8. It shows an increased extinction cross section with ka.

To give a more intuitive view of the range of changes in the extinction cross section, we choose three krill species: *Thysanoessa sp.* (10–25 mm), *M. norvegica* (20–50 mm), and *E. superba* (30–70 mm) as our target species. The frequency range is 30–420 kHz, which covers most of the commercially available acoustic systems for krill surveys. Following Ref. 39, we take contrast parameters $g=1.045$ and $h=1.025$ for *Thysanoessa sp.* and $g=1.048$ and $h=1.035$ for *M. norvegica*. For *E. superba*, the measured contrast parameters $g=1.029$, $h=1.026$ are used.[21,33,40] These values are lower ends of the values given in Ref. 40 ($g=1.0357\pm0.0067$ and $h=1.0279\pm0.0024$) and are expected to provide the lowest prediction of extinction cross section. The parameter $e=7.5$ is used for all three species. Computation results based on the MB–DCM are presented in Fig. 9. The extinction cross section (in m^2) is shown in the vertical axis. The three domains in the figure refer to various krill

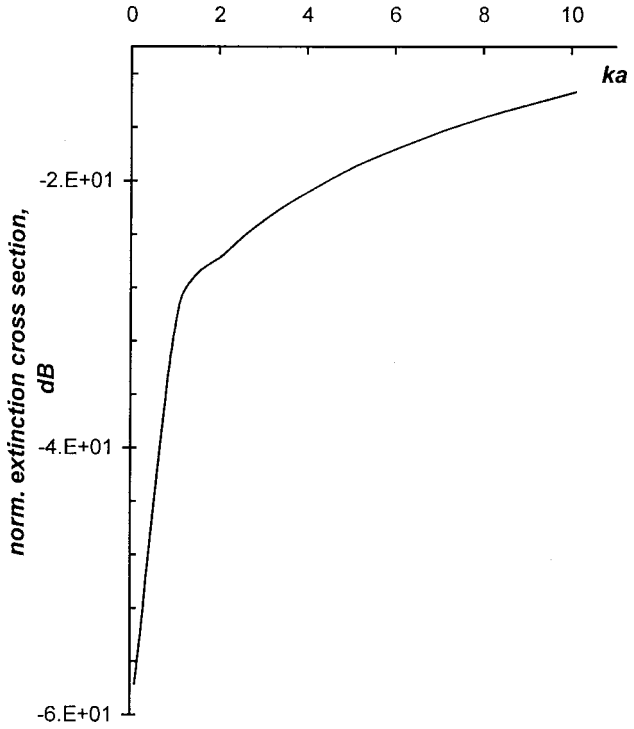


FIG. 8. The dependence of the extinction cross section on the parameter *ka*. The evaluation is based on the MB–DCM with the simulation parameters $g=h=1.04$, and the aspect ratio $e=5$. The quantity shown in the figure is $10\log_{10}(\sigma_e/l^2)$, where the extinction cross section is normalized by the square of the length of the scatterer.

species—*Thysanoessa sp*. (the domain with solid boundary), *M. norvegica* (dotted boundary), and *E. superba* (dashed boundary). The top boundary of each domain corresponds to the highest frequency used in the computations, i.e., $f_{\max}=420$ kHz, while the bottom boundary corresponds to the lowest frequency, $f_{\min}=30$ kHz. The quantities corresponding to the intermediate frequencies fall in the domain between these two lines. Figure 9 defines the range of change of the extinction cross section for various krill species at different echosounder frequencies.

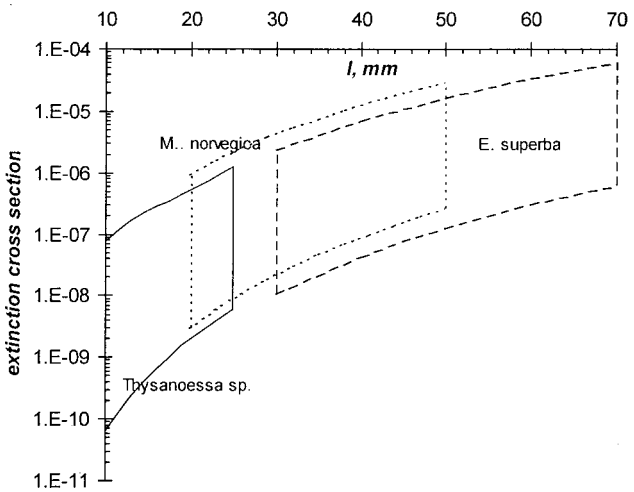


FIG. 9. The variations of the extinction cross section (in m^2) for various krill species at frequencies used in krill abundance estimation. The MB–DCM based solution is used. Three domains in the figure refer to different krill species: *Thysanoessa sp*. (solid boundary), *M. norvegica* (dotted boundary), and *E. superba* (dashed boundary). The top curve of each domain corresponds to the highest frequency $f_{\max}=420$ kHz. The bottom curve refers to the lowest frequency $f_{\min}=30$ kHz. The calculation parameters are *Thysanoessa sp.*: $g=1.045$ and $h=1.025$; *M. norvegica*: $g=1.048$ and $h=1.035$; *E. Superba*: $g=1.029$, and $h=1.026$; the aspect ratio $e=8$ for all species.

III. EXTINCTION CROSS SECTION OF AGGREGATED ZOOPLANKTON

A. Formulation

In Sec. II we have studied forward scattering by an individual target. Since sound extinction by weakly scattering zooplankton can only be important when they are densely aggregated, we need to evaluate the overall forward scattering cross section by a cloud of zooplankton. To estimate scattering by randomly aggregated zooplankton, an incoherent average over a number of distributions: size, orientation, elongation, *g*, and *h* should be used. To reduce the computation intensity in the following analysis, we will use prolate spheroids as the presumed shapes of the target species.
The averaged extinction cross section can be calculated from the following formula:

$$\langle\sigma_{\text{ext}}\rangle=\int\int\int da\,de\,dh\,dg\,d\gamma\,W_a(a)W_e(e)W_h(h)\times W_g(g)W_\gamma(\gamma)\sigma_{\text{ext}}, \tag{12}$$

where the averaging is taken over the semiminor axis *a* of the prolate spheroids, the aspect ratio $e=l/(2a)$, the sound speed and density contrasts *h* and *g*, and the orientation angle γ described in Fig. 1. In Eq. (12) the functions $W_a(a)$, $W_e(e)$, $W_h(h)$, $W_g(g)$, and $W_\gamma(\gamma)$ describe the probability density functions of the zooplankton over the parameters *a*, *e*, *h*, *g*, and γ, respectively. Without losing generality, we use the Gaussian distribution for all parameters *a*, *e*, *h*, *g*, and γ. The probability density function (PDF) of a Gaussian distribution can be expressed as

$$W_x(x)=\frac{1}{\sqrt{2\pi}S_x}\exp\left[\frac{-(x-\overline{x})^2}{2S_x^2}\right], \tag{13}$$

where *x* is one of the parameters *a*, *e*, *h*, *g* or γ, $\overline{x}$ and S_x are the mean and standard deviation of the Gaussian PDF.

Using Eqs. (1), (3), (4), (8) for the MB–DCM and Eqs. (1) and (10) for PC–DWBA, we obtain the following general formulas for the averaged extinction cross section normalized by the square of the mean of length $(\overline{l})$ of the prolate spheroids:

$$\langle\sigma_{\text{ext}}\rangle/\overline{l^2}=-(\pi\overline{\beta^2}S_\beta S_h\overline{e})^{-1}\int\int d\beta\,dh\times\exp\left(-\frac{(\beta-\overline{\beta})^2}{2S_\beta^2}\right)\exp\left(-\frac{(h-\overline{h})^2}{2S_h^2}\right)\times\text{Re}\left[\int_0^1 du\sum_0^\infty b_m(\beta,h,g,\gamma=0)\right] \tag{14}$$

for the MB–DCM approach and

$$\langle \sigma_{\rm ext}\rangle/\overline{l^2}=\frac{2\pi}{9}(k\overline{a})^2\overline{e}^{-2}\left[3\frac{S_a^4}{\overline{a}^4}+6\frac{S_a^2}{\overline{a}^2}+1\right]J_hJ_e \qquad (15)$$

for the PC–DWBA approach.

In Eq. (14), the integration variable $\beta=ka$ and $\overline{\beta}=k\overline{a}$, $S_\beta=kS_a$. Integration over the aspect ratio has been performed exactly. Taking into account the slight dependence of the forward scattering amplitude on the density contrast and the orientation angle, averaging over these parameters has been done on the basis of the method of steepest descent. In Eq. (15), the functions J_h and J_e refer to the averages over the contrast parameter h and aspect ratio e, respectively, and are described accordingly by the following integrals:

$$J_h=\frac{1}{\sqrt{2\pi}S_h}\int\frac{(h^2-1)(h-1)}{h^3}\exp\left[-\frac{(h-\overline{h})^2}{2S_h^2}\right]dh, \qquad (16)$$

$$J_e=1/((2\pi)^{1/2}S_e)\int e^{4/3}\exp\left[-\frac{(e-\overline{e})^2}{2S_e^2}\right]$$

$$=\frac{\Gamma(7/3)}{(2\pi)^{1/2}}S_e^{4/3}\exp\left(-\frac{\overline{e}^2}{4S_e^2}\right)D_{-7/3}\left(-\frac{\overline{e}}{S_e}\right), \qquad (17)$$

where the function $D_{-7/3}(x)$ is the parabolic cylinder function.[41]

Clearly, the averaged extinction cross section based on the PC–DWBA model Eq. (15) is much simpler and much more efficient computationally than that based on the MB–DCM, since the former involves only 1D integrals while the latter involves a 3D integral.

The numerical results of the averaged extinction cross section using two approaches are illustrated in Fig. 10, where $10\log_{10}(\langle\sigma_{\rm ext}\rangle/\overline{l^2})$ is plotted against the dimensionless parameter ka. The following computation parameters are used in generating Fig. 10: $S_a=0.2a$, $e=8$, $S_e=0.2e$, and $\overline{h}=1.0279$, $S_h=0.25(h-1)$. Two curves in the figure refer to the different models. The comparison shows an increasing difference between the two solutions. The discrepancy increases with ka and is about 2.5 dB at $ka=10$. In obtaining Fig. 10, the values of the semiminor axis, the aspect ratio, and the sound speed contrast used in the numerical evaluations vary between ±2 standard deviation from their corresponding mean values. To obtain the analytical expression Eq. (15), we have extended the range of each integration parameter to $\pm\infty$.

B. Examples and discussions

Based on the above results, we are now able to evaluate the influence of sound extinction on sound propagation through a layer consisting of an aggregation of randomly distributed zooplankton. The attenuation can be described as[25]

$$I(x)=I_0\exp(-xn\langle\sigma_{\rm ext}\rangle), \qquad (18)$$

where $I(x)$ denotes the acoustic intensity at x within the layer (the range along the direction of the incident wave), and $I_0=I(x=0)$. n denotes the zooplankton concentration (ind./m^3).

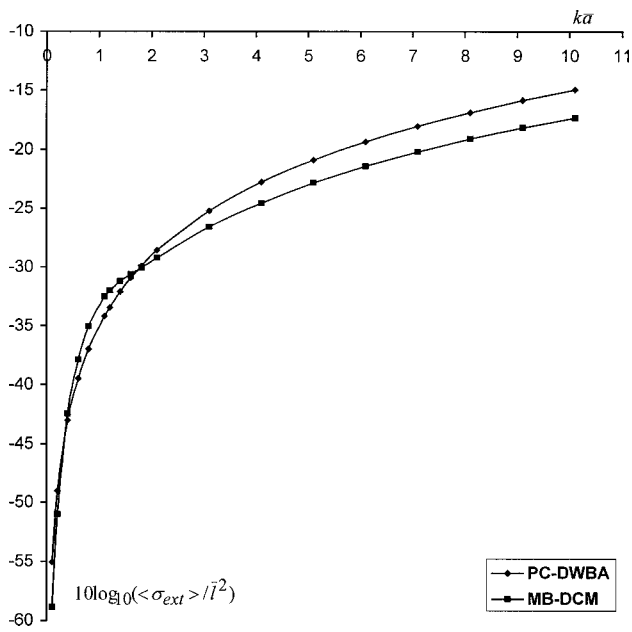


FIG. 10. The dependence of the averaged extinction cross section $10\log_{10}(\langle\sigma_e\rangle/\overline{l^2})$ on $k\overline{a}$. Two different models are used: the MB–DCM (squares) and the PC–DWBA (diamonds). The simulation parameters are $\overline{e}=8$, $\overline{h}=1.0279$, $S_a=0.2\overline{a}$, $S_e=0.2\overline{e}$, and $S_h=0.25(h-1)$. For $S_h=0.25(\overline{h}-1)$, the difference between the direct integration and the approximate solution from the method of steepest descent is less than 3.5%. However, the curve shown in the figure is based on the direct numerical integration.

The following condition

$$10\log_{10}[I(X)/I_0]<-3\ \ {\rm dB} \qquad (19)$$

is used as the criterion to define the boundary that determines qualitatively whether the sound extinction is significant or not. The cutoff threshold, -3 dB, represents 50% loss of the total intensity (energy). Using Eqs. (3), (4), and (14), we can compute the received intensity in terms of the layer thickness X, zooplankton concentration n, and dimensionless parameter $k\overline{a}$, where k is the wave number and $\overline{a}$ is the mean semiminor axis of the prolate spheroid. To make the result more general, we use a dimensionless parameter kX instead of X, volume concentration $\Phi=nV_m$, where V_m is the mean volume of individual zooplankton. Figure 11 summarizes the results of the calculations, where the same simulation parameters have been used as for Fig. 10. In the figure, the different curves correspond to the volume fractions specified in the legend. For each volume fraction, the sound extinction is considered *significant* in the domain above the curve (<-3 dB) and considered *insignificant* in the region below the curve. The estimated scattering-induced attenuation can be deduced from Fig. 11 if the acoustic frequency, mean animal size, volume fraction, and layer thickness are given. For example, assuming that the frequency$=420$ kHz, mean size $\overline{a}=2.5$ mm, volume fraction$=0.01$, layer thickness$=10$ m, and sound speed in water$=1500$ m/s, we have $ka=4.40$, and $kX=1.76\times10^4$. The corresponding location to the specified kX and $k\overline{a}$ on the $kX-k\overline{a}$ plane for such an example is marked with an open circle, which is above the $\Phi=0.01$ curve, indicating that the scattering-induced attenuation is greater than 3 dB.

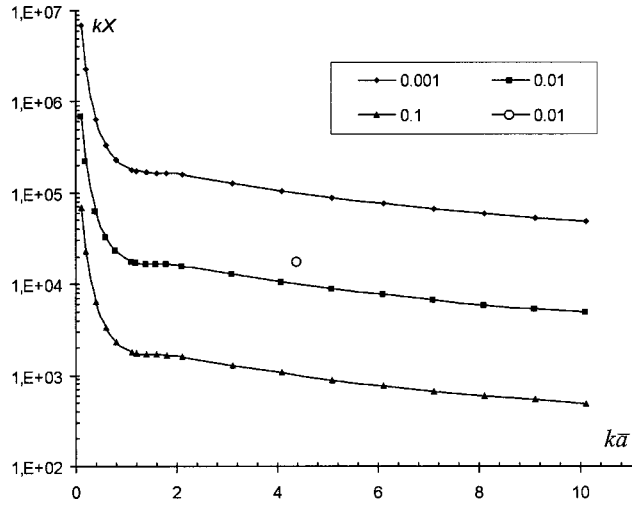


FIG. 11. Representation of the intensity reduction due to the scattering-induced attenuation on the $kX-k\bar{a}$ plane, where k is the acoustic wave number, X is the layer thickness, and $\bar{a}$ is the mean animal size (semiminor axis). Curves corresponding to the condition $10\log_{10}[I(X)/I_0]=-3$ dB are presented. Various curves refer to different volume fractions. The values of volume fraction are specified in the legend. For each volume fraction, the sound extinction is significant in the domain above the corresponding curve and can be omitted below the curve. The MB–DCM based solution, Eq. (14), and the same simulation parameters as for Fig. 10 have been used in the computations. The location of the open circle ($k\bar{a}=4.40$, and $kX=1.76e4$) corresponds to the following parameters: frequency=420 kHz, mean size $\bar{a}=2.5$ mm, volume fraction=0.01, layer thickness=10 m, and sound speed in water=1500 m/s. The location on the $kX-k\bar{a}$ plane is above the $\Phi=0.01$ curve, indicating that the scattering-induced attenuation is greater than 3 dB.

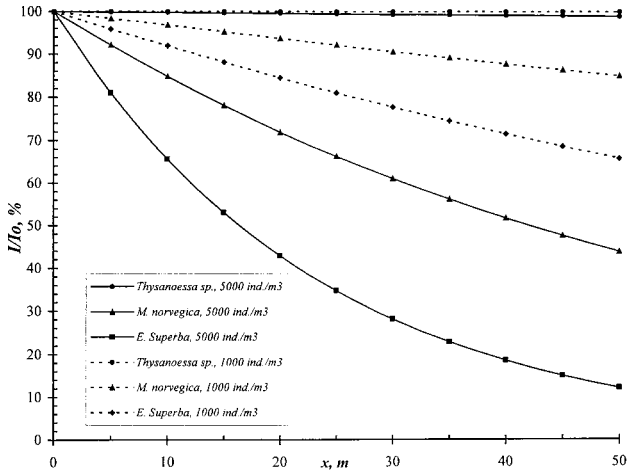


FIG. 12. Dependence of the observed intensity at the distance x in the zooplankton layer assuming the layer thickness X is layer than x. The MB–DCM based solution, Eq. (14), has been used in the computation. Different curves correspond to different krill species with different zooplankton concentrations. The curves marked by circles, triangles, and squares refer to *Thysanoessa sp.*, *M. norvegica*, and *E. Superba*, respectively. The solid and dashed curves correspond to $n=5000$ ind./m^3 and $n=1000$ ind/m^3, respectively. Simulation parameters are $\bar{l}=10$ mm, $\bar{g}=1.045$, $\bar{h}=1.025$ for *Thysanoessa sp.*; $\bar{l}=30$ mm, $\bar{g}=1.048$, $\bar{h}=1.035$ for *M. norvegica*; and $\bar{l}=45$ mm, $\bar{g}=1.029$, and $\bar{h}=1.026$ for *E. Superba*. Mean aspect ratio $\bar{e}=8$ and frequency $f=420$ kHz are used for all cases.

To obtain an alternative representation of the attenuation due to krill, we compute the scattering-induced attenuation using Eq. (18) for various krill species. The dependence of the observed intensity (%) on the distance x is demonstrated in Fig. 12. As in Fig. 9, three different krill species are used with two different zooplankton concentrations. The curves marked by circles, triangles, and squares correspond to *Thysanoessa sp.*, *M. norvegica*, and *E. Superba*, respectively. The solid and dashed curves refer to the concentrations $n=5000\,ind./m^3$ and $n=1000\,ind./m^3$, respectively. The MB–DCM based solution, Eq. (14), is employed. The simulation parameters are $\bar{l}=10$ mm, $\bar{g}=1.045$, $\bar{h}=1.025$ for *Thysanoessa sp.*, and $\bar{l}=30$ mm, $\bar{g}=1.048$, $\bar{h}=1.035$ for *M. norvegica*.[20,39] For *E. Superba*, we used $\bar{l}=45$ mm, $\bar{g}=1.029$, and $\bar{h}=1.026$.[21,33,40] The parameter $\bar{e}=8$ is used for all cases. The calculations are made for sound frequency 420 kHz.

The figure demonstrates the significant influence of extinction for larger krill species *M. norvegica* and *E. Superba* with higher concentrations. For $n=5000$ ind./m^3 and at $x=5$ m, the reductions in intensity are about 8% and 19% for *M. norvegica*, and *E. Superba*, respectively. The reductions will increase to 22% and 47%, respectively at $x=15$ m. However, for small organisms (*Thysanoessa sp.*), even with $n=5000$ ind./m^3, the extinction effect can be neglected.

IV. CONCLUSIONS

In this paper a systematic analysis has been presented to provide more information associated with the sound extinction phenomenon in the presence of zooplankton. The methodological part of the analysis includes several components.

(1) Solutions based on both the MB–DCM and PC–DWBA are presented to describe the extinction cross section of zooplankton. The MB–DCM based analytical expressions of the forward scattering amplitude for various geometrical shapes: straight cylinders, bent cylinders, and prolate spheroids are provided explicitly.

(2) Accuracy of the MB–DCM based and PC–DWBA based solutions is studied. The comparison of the MB–DCM based backward and forward scattering amplitudes with the exact solutions shows remarkable agreement. Although the PC–DWBA based solutions are not as accurate as those based on the MB–DCM, they do provide more physical insight into the scattering problem and show a reasonable agreement with the exact solutions.

(3) Analytical expressions for the extinction cross section averaged over size, aspect ratio, sound speed contrast, density contrast, and target orientation are obtained for both the MB–DCM and PC–DWBA approaches. The averaged extinction cross section based on the PC–DWBA model, Eq. (15), is much simpler (only two 1D integrations) than Eq. (14) based on the MB–DCM (a 3D integration) and is much more efficient computationally.

In addition to the methodological part, a systematic analysis of the model sensitivities for forward scattering has been presented. Simulations based on the MB–DCM suggest that the shape of the scatterer more significantly influence backscattering than forward scattering. This result is consistent with the solutions based on the PC–DWBA. The analysis demonstrates that, similar to backscattering, absorption in the individual tissue for the forward scattering is small and does not have much impact on the scattering amplitude.

Simulations based on both MB–DCM and PC–DWBA models also show a monotonic increase of the normalized extinction cross section with the dimensionless parameter *ka*, or more specifically, a $(ka)^2$ dependence.

Moreover, scattering-induced attenuation is studied for sound propagation through a layer of densely aggregated zooplankton. The influence region where the acoustic intensity is reduced more than 50% due to sound extinction by zooplankton can be defined in terms of various parameters, which include the product of the acoustic wave number and layer thickness, *kX*, the product of the acoustic wave number and the mean cylindrical radius, *ka*, and the volume fraction, Φ. It is also demonstrated that, for larger krill species, sound extinction by densely aggregated zooplankton can cause a significant reduction in the measured acoustic intensity (as much as an 85% reduction in acoustic intensity for *E. superba* with $n = 5000$ ind./m^3 over a range of 50 m).

ACKNOWLEDGMENTS

This work is partially sponsored by the Institute of Oceanology, Polish Academy of Sciences (the sponsor program 2.7), and the National Science Foundation of the United States, Grant No. OCE-9730680.

[1] K. G. Foote, E. Ona, and R. Toresen, ''Determining the extinction cross section of aggregating fish,'' J. Acoust. Soc. Am. **91**, 1983–1989 (1992).
[2] I. E. Davies, ''Attenuation of sound by schooled anchovies,'' J. Acoust. Soc. Am. **54**, 213–217 (1973).
[3] Ó. Diachok, ''Effects of absorptivity due to fish on transmission loss in shallow water,'' J. Acoust. Soc. Am. **105**, 2107–2128 (1999).
[4] M. Furusawa, K. Ishii, and Y. Miyanohana, ''Attenuation of sound by schooling fish,'' J. Acoust. Soc. Am. **92**, 987–994 (1992).
[5] T. Hashimoto, ''Characteristics of ultrasonic wave transmitted horizontally in the water,'' Rep. Fish. Boat. Lab. **2**, 55–58 (1955) (in Japanese).
[6] K. Ishii, M. Furasawa, and Y. Miyanohana, ''Measurements of attenuation of sound by schooling fish,'' Tech. Rep. Nat. Res. Inst. Fish. Eng. (Fishing Boat and Instrum.) **4**, 73–94 (1983) (in Japanese).
[7] K. Ishii, M. Furasawa, and Y. Miyanohana, ''Measurements of attenuation of sound by schooling fish,'' Proc. Acoust. Soc. Jpn. **1985**, 715–716 (1985) (in Japanese).
[8] K. G. Foote, ''Analysis of empirical observations on the scattering of sound by encaged aggregation of fish,'' Fiskeridir. Skr., Ser. Havunders. **16**, 423–456 (1978).
[9] M. Furusawa, K. Ishii, Y. Miyanohana, and Y. Maniwa, ''Experimental investigation of an acoustical method to estimate fish abundance using culture nets,'' Jpn. J. Appl. Phys., Suppl. **23**, 101–103 (1984).
[10] F. Armstrong, E. J. Simmonds, and D. N. MacLennan, ''Sound losses through aggregations of fish,'' Proc. Institute of Acoustics **11**, 35–43 (1989).
[11] D. N. MacLennan, F. Armstrong, and E. J. Simmonds, ''Further observations on the attenuation of sound by aggregations of fish,'' Proc. Institute of Acoustics **12**, 99–106 (1990).
[12] K. Olsen, ''Sound attenuation within schools of herring,'' ICES CM B **44**, 15 (1986).
[13] R. Toresen, ''Absorption of acoustic energy in dense herring schools studied by the attenuation in the bottom echo signal,'' Fish. Res. **10**, 317–327 (1991).
[14] Z. Ye, ''On acoustic attenuation by swimbladder fish,'' J. Acoust. Soc. Am. **100**, 669–672 (1996).
[15] D. Chu and Z. Ye, ''A phase-compensated DWBA representation of the bistatic scattering. Weakly scattering objects: application to zooplankton,'' J. Acoust. Soc. Am. **106**, 1732–1743 (1999).
[16] N. Gorska, ''Evaluation of sound extinction and echo interference in densely aggregated zooplankton,'' Oceanologia **42**, 1–20 (2000).
[17] V. Siegal, ''Krill (Euphausiacea) demography and variability in abundance and distribution,'' Can. J. Fish. Aquat. Sci. **57**, 151–167 (2000).
[18] J. Mauchline, *Biology of Euphausiids, Advances in Marine Biology* (Academic, New York, 1980), pp. 524–526, 613–615.
[19] W. M. Hamner and P. P. Hamner, ''Behavior of Antarctic krill (*Euphausia superba*): schooling, forging, and antipredatory behavior,'' Can. J. Fish. Aquat. Sci. **57**, 192–202 (2000).
[20] J. W. Kogeler, S. Falk-Petersen, A. Kristensen, F. Pettersen, and J. Dalen, ''Density- and sound speed contrasts in sub-Arctic zooplankton,'' Polar Biol. **7**, 231–235 (1987).
[21] K. G. Foote, ''Speed of sound in *Euphausiid Superba*,'' J. Acoust. Soc. Am. **87**, 1405–1408 (1990).
[22] D. Chu, P. H. Wiebe, and N. Copley, ''Inference of material properties of zooplankton from acoustic and resistivity measurements,'' ICES J. Mar. Sci. **57**, 1128–1142.
[23] G. J. Gruber and R. Meister, ''Ultrasonic attenuation in water containing brine shrimp in suspension,'' J. Acoust. Soc. Am. **33**, 733–740 (1961).
[24] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978), Vol. 1.
[25] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, St. Louis, San Francisco, Toronto, London, Sydney, 1968).
[26] D. T. DiPerna and T. K. Stanton, ''Sound scattering by cylinders of noncircular cross section: A conformal mapping approach,'' J. Acoust. Soc. Am. **96**, 3064–3079 (1994).
[27] C. Partridge and E. R. Smith, ''Acoustic scattering from bodies: Range of validity of the deformed cylinder method,'' J. Acoust. Soc. Am. **97**, 784–795 (1995).
[28] Z. Ye, ''A novel approach to sound scattering by cylinder of finite length,'' J. Acoust. Soc. Am. **102**, 877–884 (1997).
[29] Z. Ye, E. Hoskinson, E. Dewey, R. K. Ding, and D. M. Farmer, ''A method for acoustic scattering by slender bodies. I. Theory and verification,'' J. Acoust. Soc. Am. **102**, 1964–1976 (1997).
[30] T. K. Stanton, ''Sound scattering by cylinder of finite length. III. Deformed cylinders,'' J. Acoust. Soc. Am. **86**, 691–705 (1989).
[31] T. K. Stanton, D. Chu, and P. H. Wiebe, ''Sound scattering by several zooplankton groups. Scattering models,'' J. Acoust. Soc. Am. **103**, 236–253 (1998).
[32] T. K. Stanton, D. Chu, P. H. Wiebe, and C. S. Clay, ''Average echoes from randomly oriented random-length finite cylinders: Zooplankton models,'' J. Acoust. Soc. Am. **94**, 3463–3472 (1993).
[33] D. Chu, K. G. Foote, and T. K. Stanton, ''Further analysis of target strength of Antarctic krill at 38 and 120 kHz: Comparison with deformed cylinder model and inference of orientation distribution,'' J. Acoust. Soc. Am. **93**, 2985–2988 (1993).
[34] C. Yeh, ''Scattering of acoustic waves by a penetrable prolate spheroid,'' J. Acoust. Soc. Am. **42**, 518–521 (1967).
[35] *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*, edited by M. Abramowitz and A. Stegun (Dover, New York, 1972).
[36] G. C. Gaunaurd, ''Sonar cross sections of bodies partially insonified by finite sound beams,'' IEEE J. Ocean. Eng. **OE-10**, 213–230 (1985).
[37] R. H. Vogt, L. Flax, L. R. Dragonette, and W. G. Neubauer, ''Monostatic reflection of a plane wave from an absorbing sphere,'' J. Acoust. Soc. Am. **57**, 558–561 (1975).
[38] K. G. Foote, ''Absorption term in time-varied-gain functions,'' Fiskeridir. Skr., Ser. Havunders. **17**, 191–213 (1981).
[39] A. Kristensen and J. Dalen, ''Acoustic estimation of size distribution and abundance of zooplankton,'' J. Acoust. Soc. Am. **80**, 601–611 (1986).
[40] K. G. Foote, I. Everson, J. L. Watkins, and D. G. Bone, ''Target strength of Antarctic krill (Euphasia superba) at 38 and 120 kHz,'' J. Acoust. Soc. Am. **87**, 16–24 (1990).
[41] I. S. Gradshteyn and I. M. Ryzhyk, *Table of Integrals, Series, and Products*, corrected and enlarged edition (Academic, New York, 1980).

Visualization of pass-by noise by means of moving frame acoustic holography

Soon-Hong Park[a] and Yang-Hann Kim[b]
Center for Noise and Vibration Control (NOVIC), Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Science town, Taejon 305-701, Korea

(Received 15 March 2000; accepted for publication 27 July 2001)

The noise generated by pass-by test (ISO 362) was visualized. The moving frame acoustic holography was improved to visualize the pass-by noise and predict its level. The proposed method allowed us to visualize tire and engine noise generated by pass-by test based on the following assumption; the noise can be assumed to be quasistationary. This is first because the speed change during the period of our interest is negligible and second because the frequency change of the noise is also negligible. The proposed method was verified by a controlled loud speaker experiment. Effects of running condition, e.g., accelerating according to ISO 362, cruising at constant speed, and coasting down, on the radiated noise were also visualized. The visualized results show where the tire noise is generated and how it propagates. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1404976]

PACS numbers: 43.20.−f, 43.35.Sx, 43.60.Sx [ANN]

I. INTRODUCTION

Vehicle pass-by noise level is one of major concern to car manufacturers because the maximum noise level is strictly controlled by international regulations. Several measuring standards, e.g., ISO 362[1] and SAE J1470[2] specify methods of measuring the maximum noise level during the pass-by test. Being able to see where noise is generated and how it propagates would be very helpful to reduce the pass-by noise. This paper addresses the visualization of noise generated by the pass-by test.

Line array methods[3–16] have been widely used for localizing the noise sources of high-speed trains and moving vehicles. The linear, the cross-shaped, and the X-shaped arrays have been employed. However, the methods have inherent drawbacks. The methods only find the locations of noise sources because they estimate beam forming power on the assumed source plane. On the other hand, the nearfield acoustic holography (NAH)[17] provides all useful acoustic variables in the entire volume of interest. However, in order to apply NAH to a moving noise source, a plane array of microphones has to be attached to the vehicle.[18] The complexity and cost of this system limit the practical applicability of NAH to moving noise sources.

Moving frame acoustic holography (MFAH)[19–21] combines the simplicity of the conventional line array methods and the ability of NAH. MFAH uses a sweeping line array of microphones. The relative motion between the noise source and the line array enables us to continuously sweep a sound field and produces the hologram of the scanned plane. When a line array of microphones is standing on the ground, this method enables us to visualize the noise generated by a moving source based on NAH. This is the advantage of MFAH. Effects of moving noise sources on an obtained hologram can be neglected for low Mach number.[21] MFAH can be applied not only to a tonal component,[19] but also to a coherent band-limited one.[20]

Although noise generated by a rolling tire on a dynamometer can be visualized by means of time domain acoustic holography (TDH),[22] the demand to see or study the noise generated in the general situation of which a vehicle moves on a road never diminishes. This paper reports a method of visualization of pass-by noise. The operating condition of a vehicle must follow the way of which is defined in the pass-by noise measurement (ISO 362). MFAH cannot directly be applied to general transient source signals because the method assumes the hologram to be stationary in time. We propose a modified MFAH method so that it can be applied to a transient signal. In fact, this paper assumes the transient signal to be the sum of piecewise stationary signal. The applicability of the proposed method was verified not only by performing a controlled loud speaker experiment but also by comparing with the commercial STSF (spatial transformation of sound fields) system.[23] Then, tire and engine noise during the pass-by test was visualized based on the proposed method. The period of the measurement was less than 0.5 second. The speed change is shown to be negligible during the period. The frequency change during the period was small enough to regard the radiated sound as a quasistationary. The sound pressure and the intensity distribution of tire and engine noise during the pass-by test were obtained. The effect of running condition, e.g., cruising with constant speed, accelerating according to ISO 362, and coasting down conditions on the radiated sound were visualized.

II. THEORETICAL BACKGROUND

A. Moving frame acoustic holography (MFAH) for pass-by noise

Moving frame acoustic holography (MFAH), which is originally proposed to increase the aperture size and the spatial resolution of a hologram, allows a line array of micro-

[a]Present address: Department of Rocket Structure and Materials, Space Technology R&D Division, Korea Aerospace Research Institute (KARI), Science town, Taejon 305-600, Korea; electronic mail: shpark@kari.re.kr
[b]Electronic mail: yhkim@mail.kaist.ac.kr

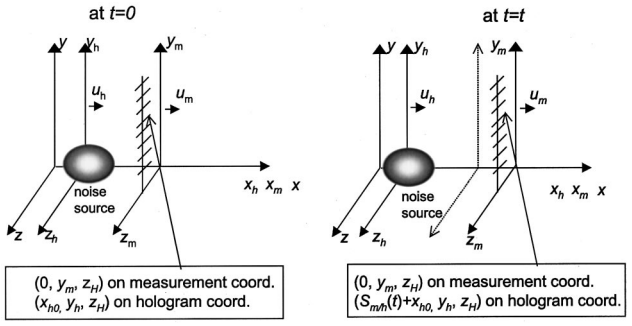


FIG. 1. Three coordinate systems employed in moving frame acoustic holography and the relative coordinate transformation. The left-hand figure indicates the coordinate systems at $t=0$. The coordinate $(0,y_m,z_H)$ on the measurement coordinate system corresponds to (x_{h0},y_h,z_H) on the hologram coordinate system. The right-hand figure shows the coordinate systems at $t=t$. Notice that the coordinate $(0,y_m,z_H)$ on the measurement coordinate system corresponds to $(S_{m/h}(t)+x_{h0}+y_h,z_H)$ on the hologram coordinate system due to the relative motion. The dashed coordinate system denotes the measurement coordinate system at $t=0$.

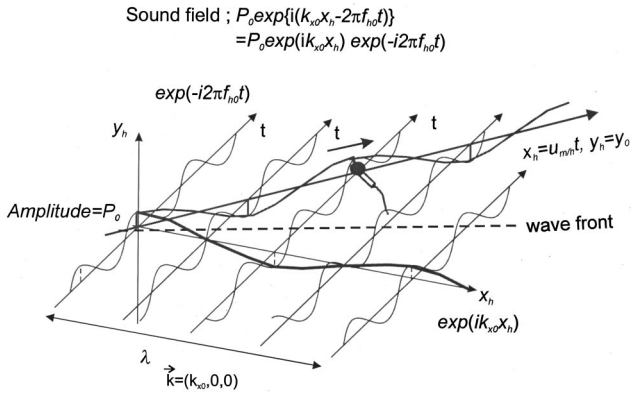


FIG. 2. Basic idea of MFAH. The plane wave sound field whose magnitude is P_0 is drawn in space and time. Notice that a sound field can be expressed as the multiplication of space and time. The frequency and the wave vector of the plane wave are f_{h0} and $(k_{x0},0,0)$. The hologram of this sound field is $P_0e^{ik_{x0}x_h}$. This can be reconstructed from the measured signal (by a moving microphone).

phones continuously sweeps a sound field. The relative motion between the line array and the noise source enables us to measure both temporal and spatial information of the sound field simultaneously. However, this induces a Doppler effect to the measured time signal. MFAH provides the method to reconstruct spatial information (sound field) from the time signal so that the hologram of the scanned plane can be obtained. The transformation introduces three coordinate systems (Fig. 1). The reference coordinate $((x,y,z))$ system is fixed to the ground (or acoustic medium). The measurement coordinate $[(x_m,y_m,z_m)$, fixed to a line array of microphones] and the hologram coordinate $[(x_h,y_h,z_h)$, fixed to the noise sources] systems are in relative motion (in x direction) with the relative velocity, $u_{m/h}(t)=u_m(t)-u_h(t)$ (see Fig. 1). We assume that the rate of change of relative velocity $u_{m/h}(t)$ is negligible. The measured sound pressure by a line array of microphones ($x_m=0$, $z_m=z_H$) can be expressed on the hologram coordinate,

$$p_m(0,y_m,z_H;t)=p_h(S_{m/h}(t),y_h,z_H;t), \tag{1}$$

where subscript m and h denote the pressure on the measurement coordinate and that on the hologram coordinate. The relative coordinate transformation,

$$x_h=S_{m/h}(t)+x_m, \quad y_h=y_m \quad \text{and} \quad z_h=z_m, \tag{2}$$

where $S_{m,h}(t)=\int_0^t u_{m/h}(\tau)d\tau$, is used in Eq. (1). When the noise source moves and the line array is fixed $[u_{m/h}(t)=-u_h(t)]$, the wave front distribution is changed due to the source motion. The wave fronts in the forward direction are closer together than they would be if the source were stationary. In the backward direction, the wave fronts are farther apart. This moving effect causes errors on the obtained hologram. However, the errors are small enough to be neglected if the speed of noise sources is much less than the speed of sound ($M<0.1$). See Ref. 21 for details. Since the speed of vehicle during pass-by test is the case therefore, MFAH could be applied to analyze the pass-by noise.

Let us start with the simplest case, so as to examine the possibility to extend MFAH to the case of our interest. Consider a moving microphone continuously scans the plane wave sound field whose magnitude is P_0. We assume that the frequency and the wave vector of the plane wave are f_{h0} and $(k_{x0},0,0)$. Then the plane wave on a plane $z_h=z_H$ can be written as (Fig. 2)

$$\begin{aligned} p_h(x_h,y_h,z_H;t)&=P_0e^{i(k_{x0}x_h-2\pi f_{h0}t)}\\ &=P_0e^{ik_{x0}x_h}e^{-i2\pi f_{h0}t}. \end{aligned} \tag{3}$$

Notice that the hologram of this sound field is $P_0e^{ik_{x0}x_h}$, which represents the spatial distribution of the sound field. This equation indicates that a sound field can be expressed as the multiplication of spatial ($P_0e^{ik_{x0}x_h}$) and temporal information ($e^{-i2\pi f_{h0}t}$) in complex domain. When the microphone moves in x direction with speed $u_{m/h}(t)=u_m(t)$, the measured pressure (time signal) preserves this property, but in different form, that is, (Fig. 2)

$$\begin{aligned} p_m(0,y_m,z_H;t)&=p_h(S_{m/h}(t),y_h,z_H;t)\\ &=P_0e^{ik_{x0}S_{m/h}(t)}e^{-i2\pi f_{h0}t}, \end{aligned} \tag{4}$$

where the microphone is fixed at $(0,y_m,z_H)$ on the measurement coordinate. This equation conveys the idea that we can obtain the hologram (the complex envelope of the signal) whenever we measure the relative displacement of microphone $x_h=S_{m/h}(t)$ and the frequency of sound field f_{h0}. The hologram can be obtained by multiplying the complex conjugate of $e^{-i2\pi f_{h0}t}$ by both sides of the equation.

This result can be extended to more general sound fields because any sound field can be expressed as the superposition of plane waves. For a general but single frequency sound field, which can be written as $p_h(x_h,y_h,z_H;t)=P_h(x_h,y_h,z_H;f_{h0})\times e^{-i2\pi f_{h0}t}$, the measured signal can be expressed as

$$\begin{aligned} p_m(0,y_m,z_H;t)&=p_h(s_{m/h}(t),y_h,z_H;t)\\ &=P_h(x_h=S_{m/h}(t),y_h,z_H;f_{h0})\times e^{-i2\pi f_{h0}t}. \end{aligned} \tag{5}$$

Notice that the hologram (P_h) can be reconstructed from the complex envelope of the measured signal.[20]

When a sound field has a continuous band-limited spec-

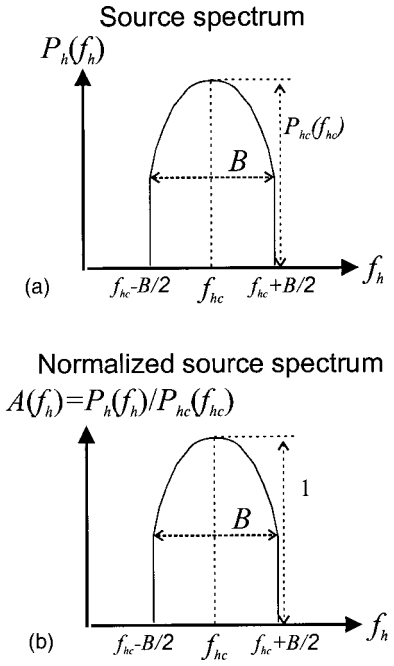


FIG. 3. Illustration of source spectrum and normalized source spectrum. The center frequency and bandwidth are denoted by f_{hc} and B. $A(f_h)$ represents the source spectrum normalized by spectrum at center frequency, that is, $A(f_h)=P_h(x_h,y_h,z_H;f_h)/P_{hc}(x_h,y_h,z_H;f_{hc})$. In practice this can be obtained from a reference microphone fixed to the hologram coordinate.

trum, the previous result can also be applied. We can express the spectrum in terms of the center frequency and bandwidth, which are denoted by f_{hc} and B (see Fig. 3). Then, the sound field can be expressed as

$$p_h(x_h,y_h,z_H;t)=\int_{f_{h-}}^{f_{h+}} P_h(x_h,y_h,z_H;f_h)\times e^{-i2\pi f_h t}\,df_h, \quad (6)$$

where $f_{h-}=f_{hc}-B/2$ and $f_{h+}=f_{hc}+B/2$. Notice that we must apply a filter to obtain this signal in practice. If the band-limited sound field is produced by coherent sources and the bandwidth is narrow, then the sound field can be expressed as[20]

$$P_h(x_h,y_h,z_H;f_h)\cong A(f_h)P_{hc}(x_h,y_h,z_H;f_{hc}) \quad (f_{h-}\leqslant f_h\leqslant f_{h+}), \quad (7)$$

where $A(f_h)$ is the normalized source spectrum, which is defined as the ratio of a complex amplitude of sound field at f_h in the band to that of center frequency [see Fig. 3(b)]. Then Eq. (6) can be written as

$$p_h(x_h,y_h,z_H;t)\simeq P_{hc}(x_h,y_h,z_H;f_{hc})\times\int_{f_{h-}}^{f_{h+}} A(f_h)e^{-i2\pi f_h t}\,df_h. \quad (8)$$

The normalized source spectrum must be obtained from measurements. A reference microphone, fixed to noise sources at $(x_{\mathrm{ref}},y_{\mathrm{ref}},z_{\mathrm{ref}})$, provides

$$A(f_h)=P_{\mathrm{ref}}(x_{\mathrm{ref}},y_{\mathrm{ref}},z_{\mathrm{ref}};f_h)/P_{\mathrm{ref},c}(x_{\mathrm{ref}},y_{\mathrm{ref}},z_{\mathrm{ref}};f_{hc}), \quad (9)$$

where $P_{\mathrm{ref}}(x_{\mathrm{ref}},y_{\mathrm{ref}},z_{\mathrm{ref}};f_h)$ and $P_{\mathrm{ref},c}(x_{\mathrm{ref}},y_{\mathrm{ref}},z_{\mathrm{ref}};f_{hc})$ represent the reference microphone spectrum of an arbitrary frequency f_h and that of center frequency f_{hc}. Notice that

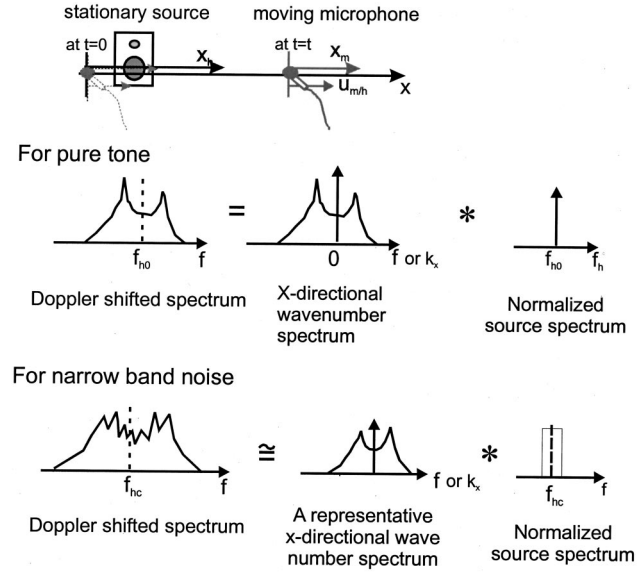


FIG. 4. Interpretation of MFAH in frequency domain and extension of MFAH for a coherent narrow band-limited noise.

$P_{\mathrm{ref},c}(x_{\mathrm{ref}},y_{\mathrm{ref}},z_{\mathrm{ref}};f_{hc})$ is constant. Therefore, Eqs. (8) and (9) give us the model of a coherent narrow band sound field.

$$\begin{aligned} p_h(x_h,y_h,z_H;t) &\cong P_{hc}(x_h,y_h,z_H;f_{hc}) \times\frac{\int_{f_{h-}}^{f_{h+}} P_{\mathrm{ref}}(x_{\mathrm{ref}},y_{\mathrm{ref}},z_{\mathrm{ref}};f_h)e^{-i2\pi f_h t}\,df_h}{P_{\mathrm{ref},c}(x_{\mathrm{ref}},y_{\mathrm{ref}},z_{\mathrm{ref}};f_{hc})} \\ &= P_{hc}(x_h,y_h,z_H;f_{hc})\frac{p_{\mathrm{ref}}(x_{\mathrm{ref}},y_{\mathrm{ref}},z_{\mathrm{ref}};t)}{P_{\mathrm{ref},c}(x_{\mathrm{ref}},y_{\mathrm{ref}},z_{\mathrm{ref}};f_{hc})}. \end{aligned} \quad (10)$$

This equation states that a coherent narrow band sound field can be modeled as the product of the representative sound field (spatial information) and the normalized source signal (temporal information). Then the previous procedure [Eqs. (4) and (5)] can also be applied to the coherent narrow band sound field. The measured pressure by the line array can be written as

$$\begin{aligned} p_m(0,y_m,z_H;t) &= p_h(S_{m/h}(t),y_h,z_H;t) \\ &= P_{hc}(x_h=S_{m/h}(t),y_h,z_H;f_{hc}) \times\frac{P_{\mathrm{ref}}(x_{\mathrm{ref}},y_{\mathrm{ref}},z_{\mathrm{ref}};t)}{P_{\mathrm{ref},c}(x_{\mathrm{ref}},y_{\mathrm{ref}},z_{\mathrm{ref}};f_{hc})}. \end{aligned} \quad (11)$$

The representative hologram (P_{hc}) of the narrow band sound field can be obtained if we measure the complex signal of the line array (p_m) and that of the reference microphone ($p_{\mathrm{ref}}/P_{\mathrm{ref},c}$) simultaneously. Figure 4 clearly demonstrates this concept in frequency domain; MFAH for a coherent narrow band noise [Eq. (11)] is compared with that of sinusoidal one [Eq. (5)].

Moving frame acoustic holography cannot directly be applied to the visualization of general transient noise sources because the spatial distribution of sound pressure on the hologram coordinate varies with respect to time. Strictly speaking, a plane array has the ability to obtain a general transient

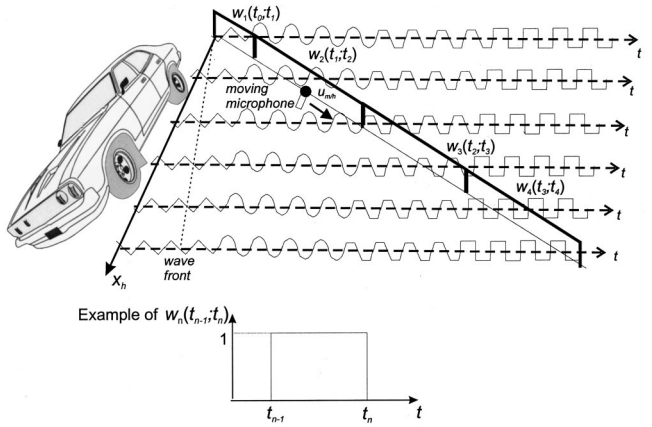


FIG. 5. Extension of MFAH for a transient noise source that can be assumed as piecewise quasistationary. Time window enables us to apply MFAH to the transient noise source. Notice that the time window can also be interpreted as a spatial window, which divides transient sound field into quasistationary ones.

hologram. The speed change of the vehicle during the pass-by test makes the sound field generated by a rolling tire transient as well. However, if the speed change is negligible during the period of our interest, then the sound field can be assumed to be stationary. We call the sound field of this kind quasistationary in time. In practice, this assumption can be made because MFAH uses the time signal less than 0.5 second in length for the visualization of pass-by noise; The speed and frequency changes are small enough to be neglected in this period of time (this will be illustrated in the next section). Then, the quasistationary sound field can be modeled as the product of a representative hologram and the source signal. This allows us to apply MFAH to the visualization of pass-by noise.

The above can be extended to a general transient noise source that can be regarded as piecewise quasistationary. We divide the measured signal by a line array into the series of time signal. This is readily done by introducing time window w_n, that is,

$$p_h(S_{m/h}(t),y_h.z_H;t)=\sum_{n=1}^{N} p_h^{(n)}(S_{m/h}(t),y_h,z_H;t_{n-1}\leqslant t\leqslant t_n), \tag{12}$$

where $p_h^{(n)}(S_{m/h}(t),y_h,z_H;t_{n-1}\leqslant t\leqslant t_n)=p_h(S_{m/h}(t),y_h,z_H;t)w_n(t_{n-1};t_n)$ is the nth Doppler shifted signal to which MFAH can directly be applied (Fig. 5). We denote the hologram of the nth Doppler shifted signal by $P_h^{(n)}(x_h^{(n-1)}\leqslant x_h\leqslant x_h^{(n)},y_h,z_H;f_h^{(n)})$ where $x_h^{(n-1)}=S_{m/h}(t_{n-1})$ and $x_h^{(n)}=S_{m/h}(t_n)$. The frequency $f_h^{(n)}$ denotes the center frequency of reference signal corresponding to the nth Doppler shifted one. Then, the hologram can be readily obtained by employing Eq. (11). After obtaining the N holograms, a representative hologram of the transient source can be obtained by adding the holograms, that is,

$$P_h(x_h,y_h,z_H)=\sum_{n=1}^{N} P_h^{(n)}(x_h^{(n-1)}\leqslant x_h\leqslant x_h^{(n)},y_h,z_H;f_h^{(n)}). \tag{13}$$

The windowed time signal reduces not only frequency resolution of the signal but also wave number resolution of the hologram. In extreme case, if the window is very short, we cannot distinguish the frequency components of our interest from other ones. Notice that the spatial resolution of this hologram is the same as what would be obtained if no time window were used to the measured time signal. It can be determined by the product of the relative velocity and the sampling period.

B. Reconstructed magnitude and the method of correction

In this section, the reconstructed magnitude by MFAH is analyzed. For a single frequency sound field , it is clear that MFAH provides the exact magnitude of the reconstructed hologram. Equations (4) and (5) illustrate this.

However, MFAH cannot provide the exact magnitude of a narrow band sound field because it reconstructs a hologram (representative hologram) of the sound field based on the model [Eq. (8)]. From Eq. (11), the magnitude of reconstructed sound field can be denoted by $|P_{h,B}|\equiv|P_{hc}(x_h,y_h,z_H;f_{hc})|$, which is the magnitude of the representative hologram at the center frequency. The true magnitude of hologram ($|P_{h,B}^{\text{true}}|$) of a narrow band sound field can be defined as

$$\begin{aligned}|P_{h,B}^{\text{true}}|&=\sqrt{\langle p_h^2(x_h,y_h,z_h;t)\rangle_{\text{av}}}\\&=\sqrt{\lim_{T\to\infty}\int_{-T/2}^{T/2} p_h^2(x_h,y_h,z_h;t)dt},\end{aligned} \tag{14}$$

where $\langle\ \rangle_{\text{av}}$ means time average. The time average of p_h^2 can be written by using Parseval theorem,

$$\begin{aligned}&\sqrt{\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} p_h^2(x_h,y_h,z_h;t)dt}\\&=\sqrt{\lim_{T\to\infty}\frac{1}{T}\int_{-\infty}^{\infty} P_h^2(x_h,y_h,z_h;t)df_h}\\&=\sqrt{\lim_{T\to\infty}\frac{1}{T}\int_{f_{h-}}^{f_{h+}} P_h^2(x_h,y_h,z_h;f_h)df_h},\end{aligned} \tag{15}$$

where P_h^2 denotes the spectral density. In discrete domain, $1/T$ will be the spectral resolution Δf_h (that is, $1/T=\Delta f_h$) and $f_h=k\Delta f_h$ (if k is integer). Therefore, Eqs. (14) and (15) can be written as

$$\begin{aligned}|P_{h,B}^{\text{true}}|&=\sqrt{\sum_k P_h^2(x_h,y_h,z_h;k\Delta f_h)(\Delta f_h)^2}\\&=\sqrt{\sum_k |P_{h,k}(x_h,y_h,z_h;k\Delta f_h)|^2},\end{aligned} \tag{16}$$

where $P_{h,k}(x_h,y_h,z_h;k\Delta f_h)\equiv P_h(x_h,y_h;z_h;k\Delta f_h)\Delta f_h$. Therefore the correction factor (a), which can be defined as $|P_{h,B}|_x a=|P_{h,B}^{\text{true}}|$ to correct the reconstructed magnitude is

$$a=\frac{|P_{h,B}^{\text{true}}|}{|P_{h,B}|}=\sqrt{\sum_k |P_{h,k}(x_h,y_h,z_h;k\Delta f_h)|^2 \Big/ |P_{hc}(x_h,y_h,z_h;f_{hc})|^2}. \tag{17}$$

In practice, the correction factor has to be obtained from the reference microphone signal,

$$a_{\text{ref}}=\sqrt{\sum_k |P_{\text{ref},k}(x_{\text{ref}},y_{\text{ref}},z_{\text{ref}};k\Delta f_h)|^2 \Big/ |P_{\text{ref}c}(x_{\text{ref}},y_{\text{ref}},z_{\text{ref}};f_{hc})|^2}. \tag{18}$$

When the sound field is a coherent narrow band one, this can be used instead of Eq. (17), that is

$$a=a_{\text{ref}}. \tag{19}$$

The proof of Eq. (19) is presented in Appendix A.

Figure 6 demonstrates that the proposed correction method is practically applicable by means of numerical simulation. Two out of phase monopoles radiate a coherent narrow band noise whose center frequency and bandwidth are 450 Hz and 30 Hz (6.7% of the center frequency). The distance between them is one wavelength [Fig. 6(a)]. A reference microphone is installed near the source [see Fig. 6(a)]. We assume that a microphone moves in x direction with speed of 10 m/s. The correction method well recovers the magnitude of true hologram [Fig. 6(b)]. Recall that this result can be applied to the case of moving sources and fixed microphones directly when the speed of noise sources is much less than the speed of sound.

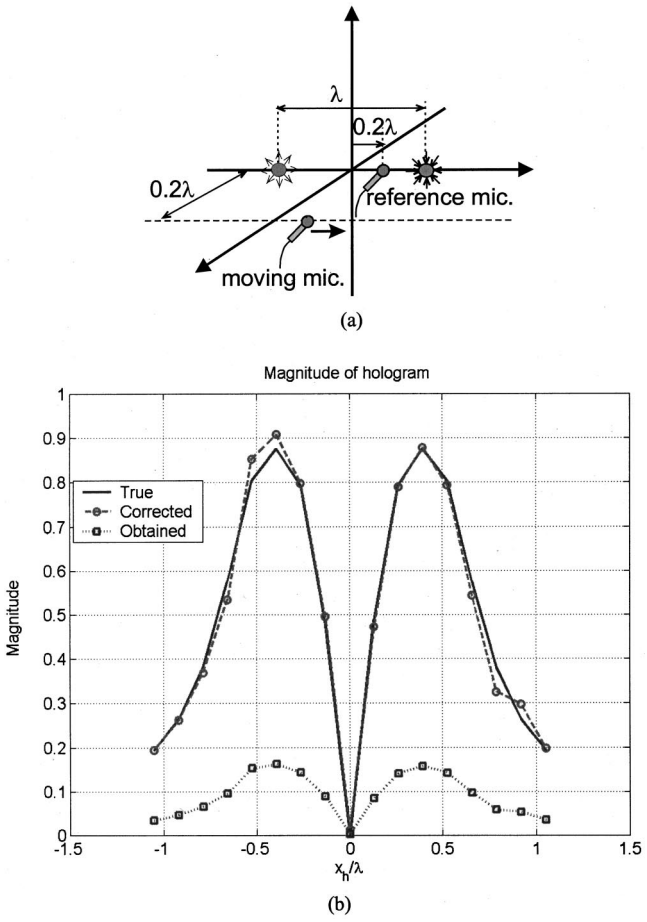


FIG. 6. The feasibility of magnitude correction factor for a coherent narrow band noise. (a) Simulation configuration. (b) Comparison of the corrected ($|P_{h,B}|\times\alpha$), the obtained ($|P_{h,B}|$), and the true ($|P_{h,B}^{\text{true}}|$) values.

III. EXPERIMENTAL RESULTS

A. Experimental setup

In order to verify the applicability of moving frame acoustic holography to pass-by noise, a set of complicated but well controlled experiment was performed. Figure 7 illustrates experimental setup. The vertical line array of 16 microphones was used to measure the Doppler shifted signal. The microphone spacing was 0.1 m so that spatial aliasing below 1 kHz can be avoided. The time signal was recorded by using a multichannel signal analyzer that can sample 32 channels simultaneously. For verification purpose, a loudspeaker unit was used to produce a controlled sound. We attached it to the right side of the vehicle.

We made special tread pattern tires that radiate narrow band noise. Two vehicles were used. One was specially de-

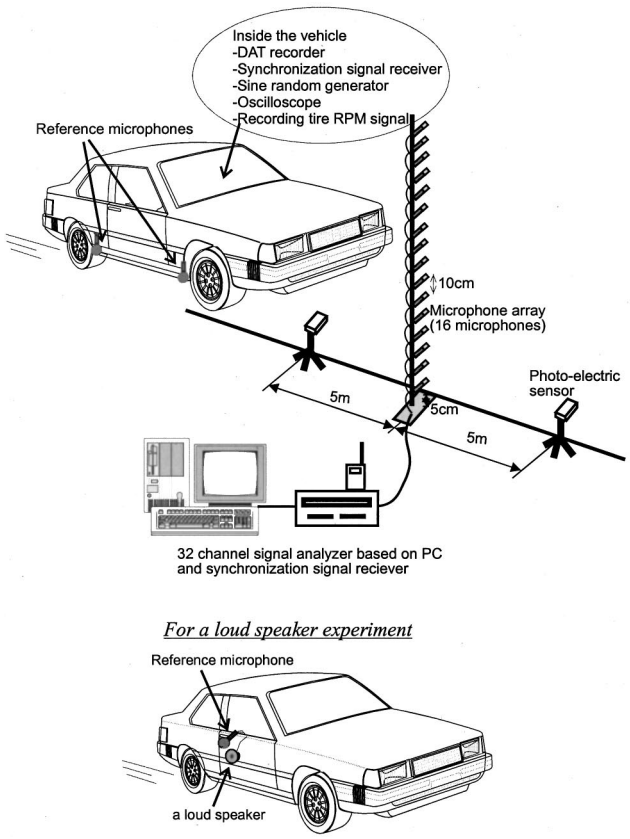
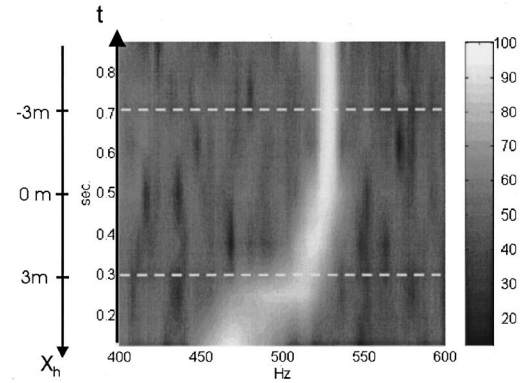
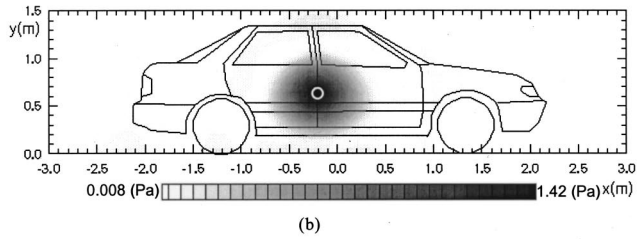


FIG. 7. Experimental setup.



Running condition-according to ISO 362 pass-by test

Pure tone varying 520 Hz to 530 Hz



Reference experiment for validation of the method

Running condition- constant speed / 452Hz pure tone

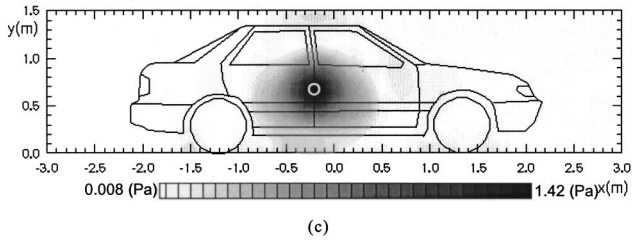
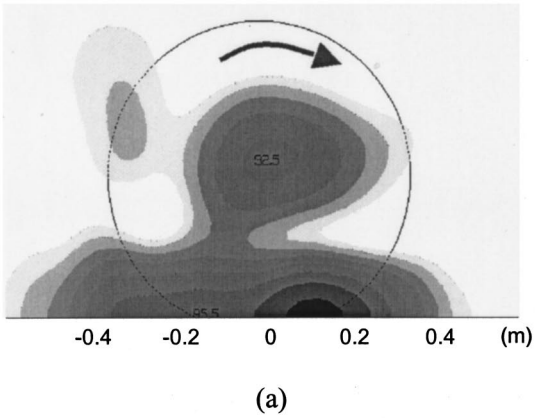


FIG. 8. Validation of MFAH for pass-by noise. A loud speaker radiates a pure tone signal varying from 520 Hz to 530 Hz. (a) Spectrogram of reference microphone signal. Dashed lines denote the period of our interest. The period is about 0.4 second. (b) Predicted sound field on a source plane. (c) Verified result for pure tone.

signed for reducing noises from the vehicle so that they would not affect the controlled sound from the loudspeaker unit and noise from the special tread pattern tires. The other was a normal one that is used for visualizing the engine noise. A tire RPM signal, which produces one period of square wave per one revolution of tire, was recorded. Running distance per one revolution of tire could be calculated by measuring time interval for vehicle to move 10 m at constant speed. Two photoelectric sensors, separated by 10 m, were used in order to measure the time interval. The tire RPM signal and two photoelectric sensors enable us to obtain the relative position between the moving vehicle and the vertical line array. They provide the relative coordinate transformation [Eq. (2)] that transforms measured signal into the pressure on the hologram coordinate.

In the case of the loudspeaker experiment, a reference microphone was installed near the loudspeaker unit to measure the source signal (or, normalized source spectrum). In the case of tire noise visualization, we installed two reference microphones near the front and rear tires (Fig. 7) to improve the observability (to avoid measuring a weak sound at the reference microphone). It is noteworthy that the proposed method requires only one reference microphone if the noise source under study emits a tonal or coherent narrow band noise. A digital audio tape (DAT) recorder was used to record



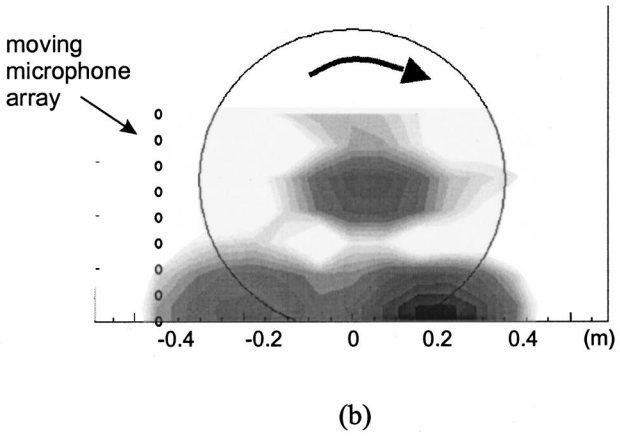
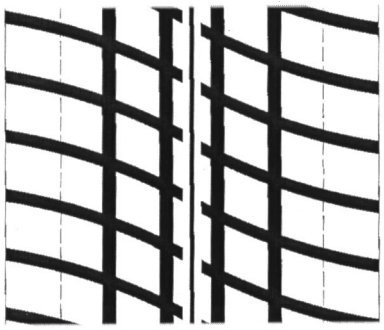


FIG. 9. Comparison between the results by STSF and MFAH. (a) Predicted sound field on a tire surface by STSF. This can be regarded as a true value. (b) Predicted sound field on a tire surface by MFAH (“O” denotes a microphone).



Belt angle 20°
Lateral groove angle 17.87°

FIG. 10. Tread pattern of the special tread pattern tire.

the reference microphone signals and the tire RPM signal. The reference microphones and the DAT recorder were installed inside the vehicle. Note that signals from the reference microphones and the array microphones must be sampled simultaneously; one synchronization signal generator and two receivers were used. The received signals were recorded by both the DAT recorder and the signal analyzer during the measurement. They allowed us to synchronize the reference and the array signals.

B. Validation of MFAH

The applicability of MFAH to moving noise sources at constant speed was verified in Ref. 20. When the vehicle accelerates according to the method of pass-by noise measurement (ISO 362), the applicability of MFAH can be verified by the following loudspeaker experiment. A loudspeaker unit attached to an accelerating vehicle radiates a sound field whose frequency varies from 520 Hz to 530 Hz [Fig. 8(a)]. Figure 8(a) shows the spectrogram of radiated sound that can be obtained from the reference microphone. Figure 8(b) shows the predicted pressure field on the source plane. Figure 8(c) shows the verified result. The verified result can be obtained when the loudspeaker unit (it radiates a 450 Hz sound field) moves with constant velocity (55 km/h). The results show a good agreement. They validate the applicability of the proposed method to pass-by noise. It is noteworthy that the sound field can be regarded as a quasistationary when the frequency change is small (about 10 Hz) during the measurement.

The applicability of MFAH to the tire noise was also verified by comparing with STSF (spatial transformation of sound fields) system,[23] which is commercial software for acoustical holographic prediction based on the step-by-step measurement. We assume that the result by the STSF system is true. Notice the STSF method uses much more informa- than the proposed method. The special tread pattern tire was installed to a dynamometer in a semianechoic chamber. it ran

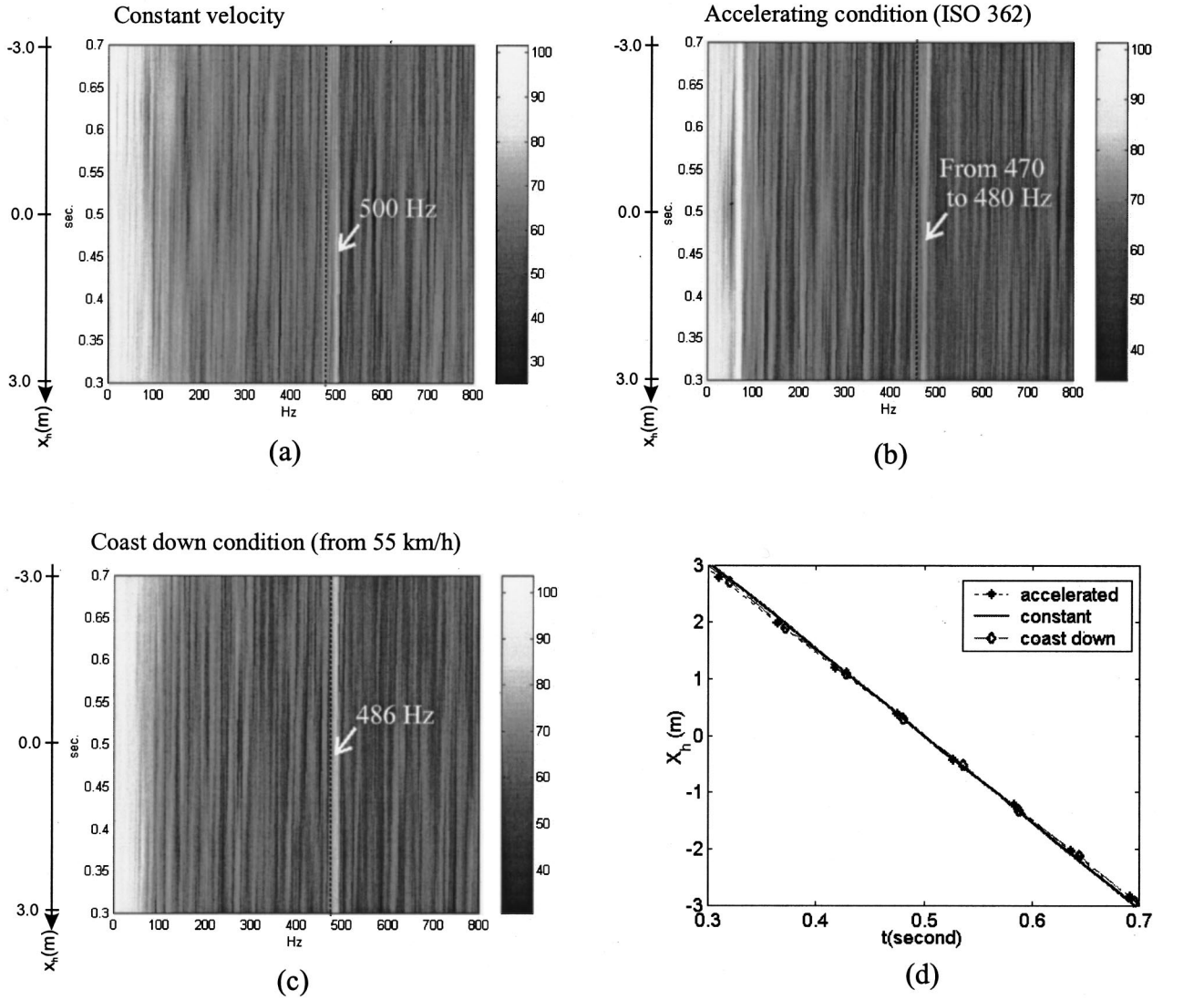


FIG. 11. Spectrogram of reference microphone for various running conditions. The lateral groove and belt angles of the tire are 17.87° and 20°. (a) Constant velocity (55 km/h). (b) Accelerating condition (from 50 km/h). (c) Coast down condition (from 55 km/h). Notice that very small frequency change occurs during the measurement period. (d) Relative position of the line array on the hologram coordinate with respect to time. Speed change is also negligible during measurement distance (6 m). The results assure that the pass-by noise during the period of our interest is quasi-stationary.

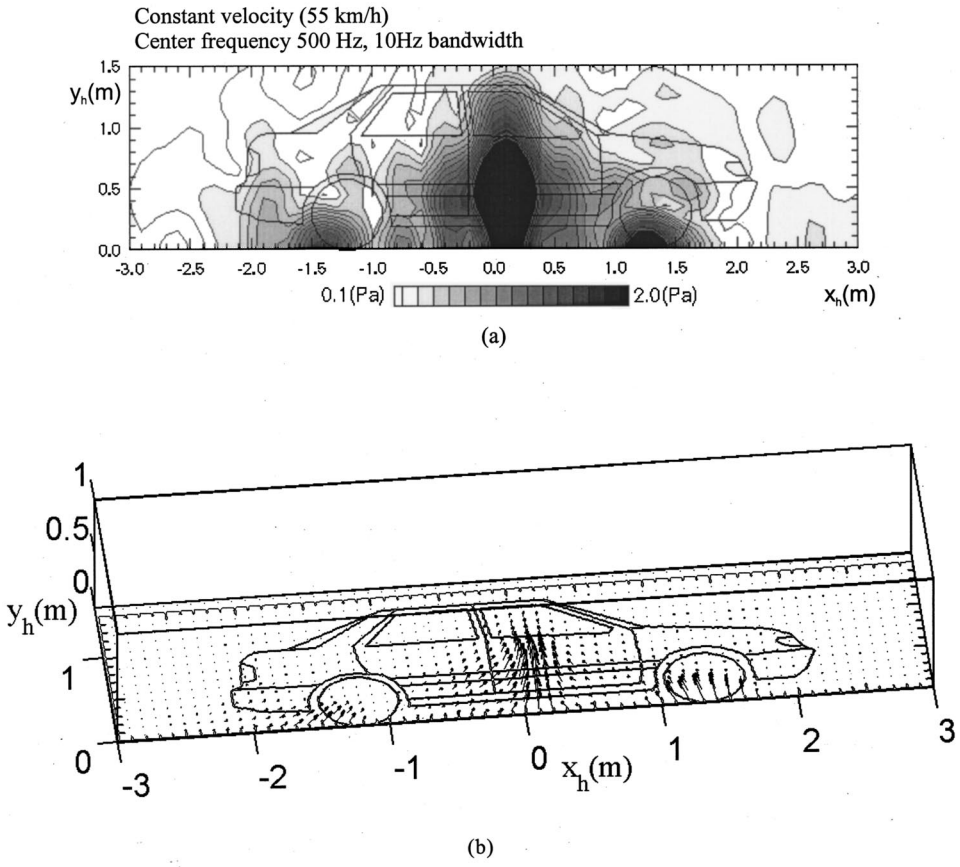


FIG. 12. The special tread pattern tire noise distribution on the assumed source plane (the surface of the vehicle, $z_h=0$) when the vehicle runs at constant velocity. (a) Sound pressure magnitude. (b) Active intensity.

at 50 km/h on the dynamometer. It radiates a narrow band noise whose center frequency is 460 Hz and 10 Hz bandwidth. In STSF method, three reference microphones were used. This enables us to identify three independent noise sources. We measured 17×14 points by using a line array that has seven microphones. A step-by-step scanning method was employed in STSF method. The microphone spacing was 6 cm. The distance between the tire surface and the hologram was 10 cm. Figure 9(a) shows the predicted pressure field on the assumed tire surface by STSF method. Very similar sound fields can be observed from the result obtained by MFAH [Fig. 9(b)]. We used a line array of nine microphones. The microphone spacing was also 6 cm. The microphone positioning system (B&K type 9665) was used to sweep the sound field. The result shows a good agreement with that by STSF (true value). Both results show that the leading edge of tire is a dominant noise source. This verified the use of MFAH to the noise generated by the special tread pattern tire. Figure 9 also implies that the noise source of the special tread pattern tire can be regarded as a coherent band-limited one.

C. Visualization of pass-by noise on a moving vehicle

We used the special tread pattern tire to visualize the tire noise generated by the pass-by test. The tire was designed to generate narrow band noise whose center frequency and bandwidth are 500 Hz and 10 Hz at 55 km/h. This enables us to see the effect of driving condition. The tread pattern of tire is shown in Fig. 10. We installed the special tread pattern tire on the right side of the vehicle (the surface of visualization). Two smooth tires, which do not have tread pattern, were installed on the other side.

Figure 11 shows spectrograms of the reference microphone according to running condition. Frequency of the tire remains constant when the vehicle runs at constant speed (55 km/h) [Fig. 11(a), compare with vertical dashed line]. When the vehicle accelerates from 50 km/h according to ISO 362, the frequency of tire noise increases about 10 Hz (from 470 Hz to 480 Hz) [Fig. 11(b)]. However, we cannot observe the frequency decrease of tire noise when the vehicle coasts down from 55 km/h [Fig. 11(c)]. Notice that the frequency variation is very small so that the tire noise can be assumed to be quasistationary. Figure 11(d) shows the relative position of array microphones on the hologram coordinate with respect to time. This illustrates that the change of vehicle speed during the period of our interest is small enough to be neglected. Figure 11 implies that the proposed method can be applied to the visualization of pass-by noise.

Figures 12–16 demonstrate the effect of driving condition on the radiation of tire noise. Figure 12 represents the pressure magnitude and the three dimensional active intensity on the assumed source plane ($z_h=0$) when the vehicle runs at constant speed (55. 25 km/h). The radiated sound from the front tire interferes with that from the rear tire at the center of the vehicle resulting in surprisingly high pressure field. Figure 13 depicts this interference pattern within one period (T) of corresponding frequency (500 Hz). When the vehicle is accelerated from 50 km/h according to ISO 362, the front tire radiates more sound than rear one does [Fig.

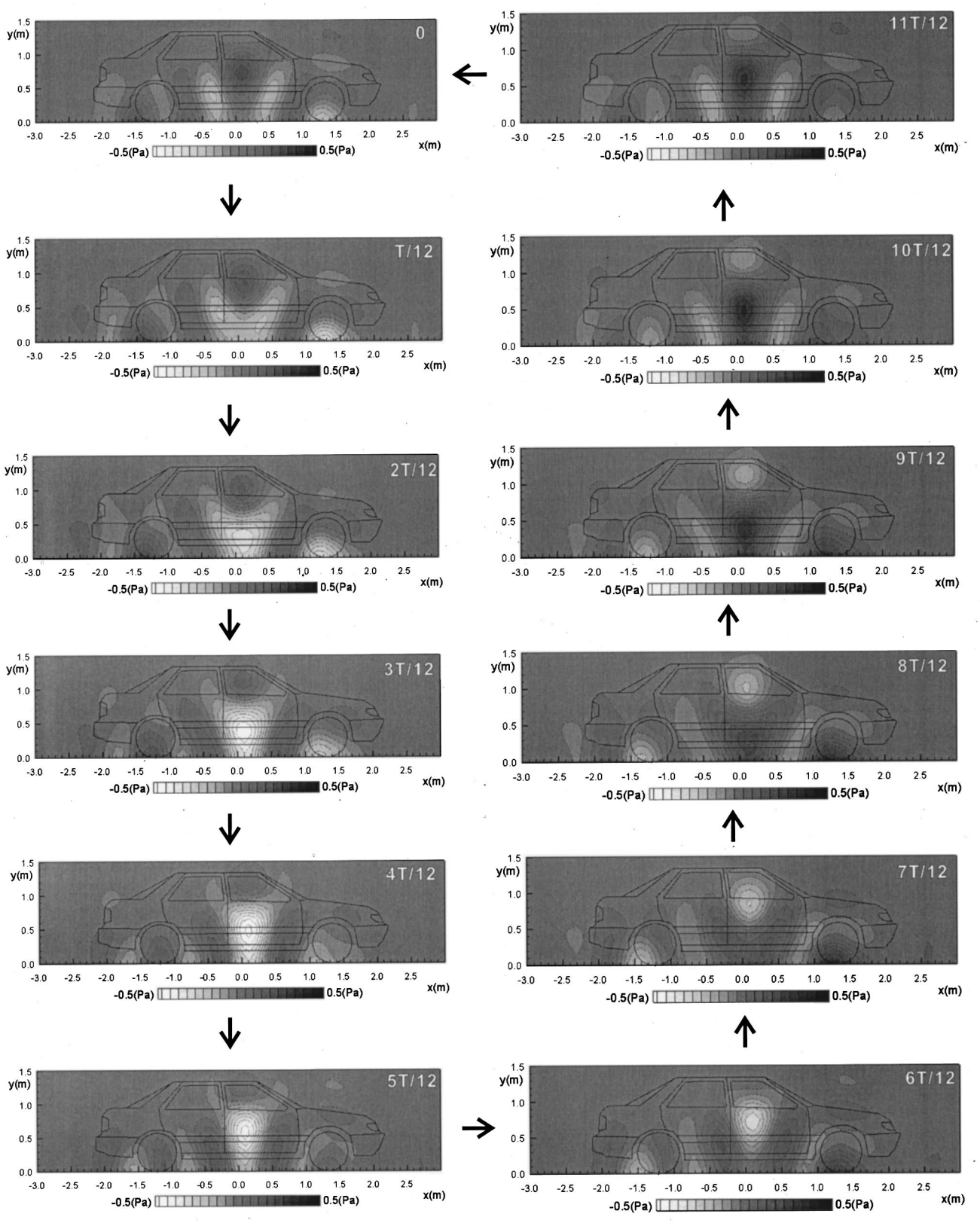


FIG. 13. The animation of sound pressure during one period (T) of corresponding frequency is presented when the vehicle runs at constant velocity. The visualization plane is $z_h = 0$. Notice that the interference between the radiated noise from the front tire and that from the rear one results in the high pressure in the center of the vehicle.

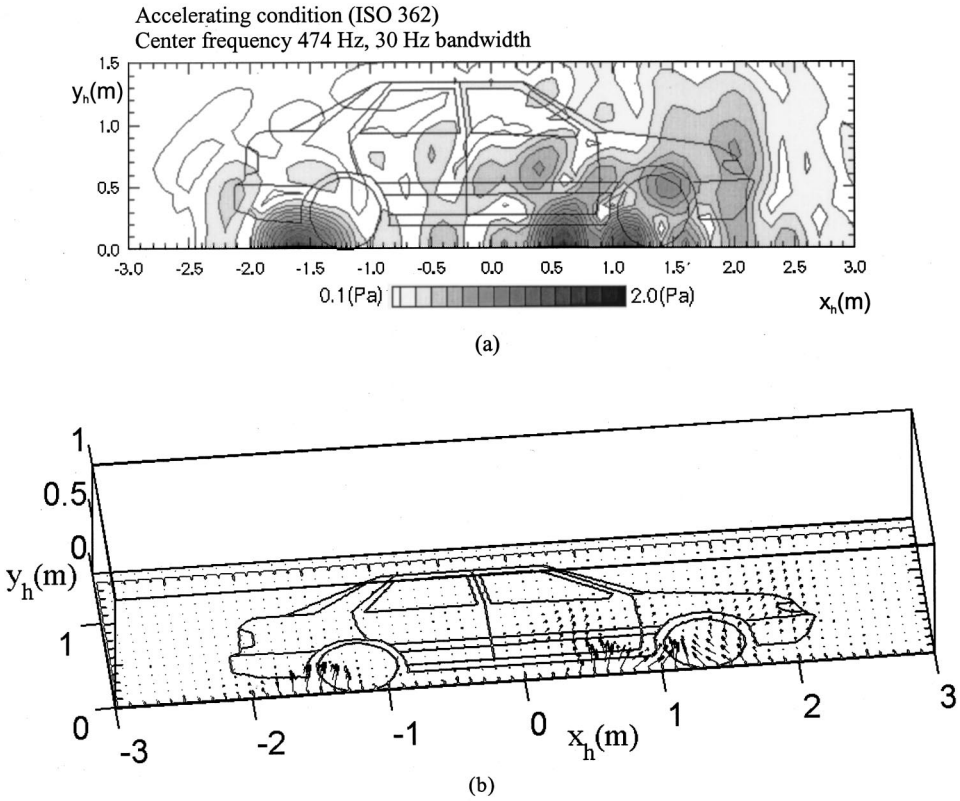


FIG. 14. The special tread pattern tire noise distribution on the assumed source plane (the surface of the vehicle, $z_h=0$) when the vehicle is accelerated from 50 km/h. (a) Sound pressure magnitude. (b) Active intensity.

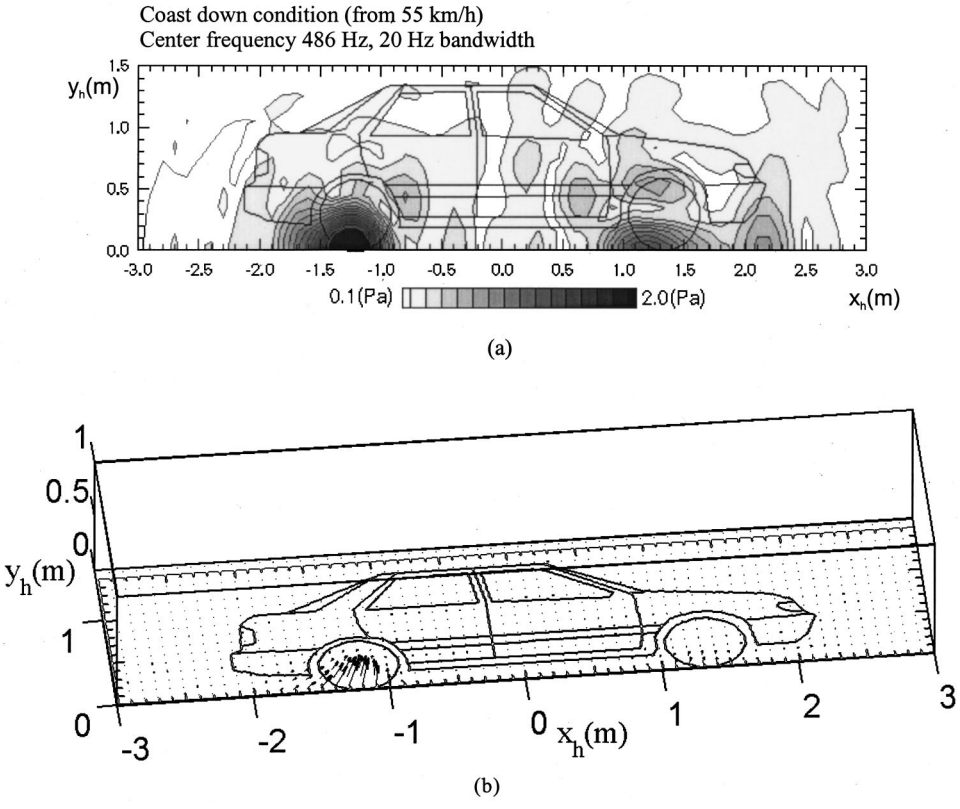


FIG. 15. The special tread pattern tire noise distribution on the assumed source plane (the surface of the vehicle, $z_h=0$) when the vehicle coasts down from 55 km/h. (a) Sound pressure magnitude. (b) Active intensity.

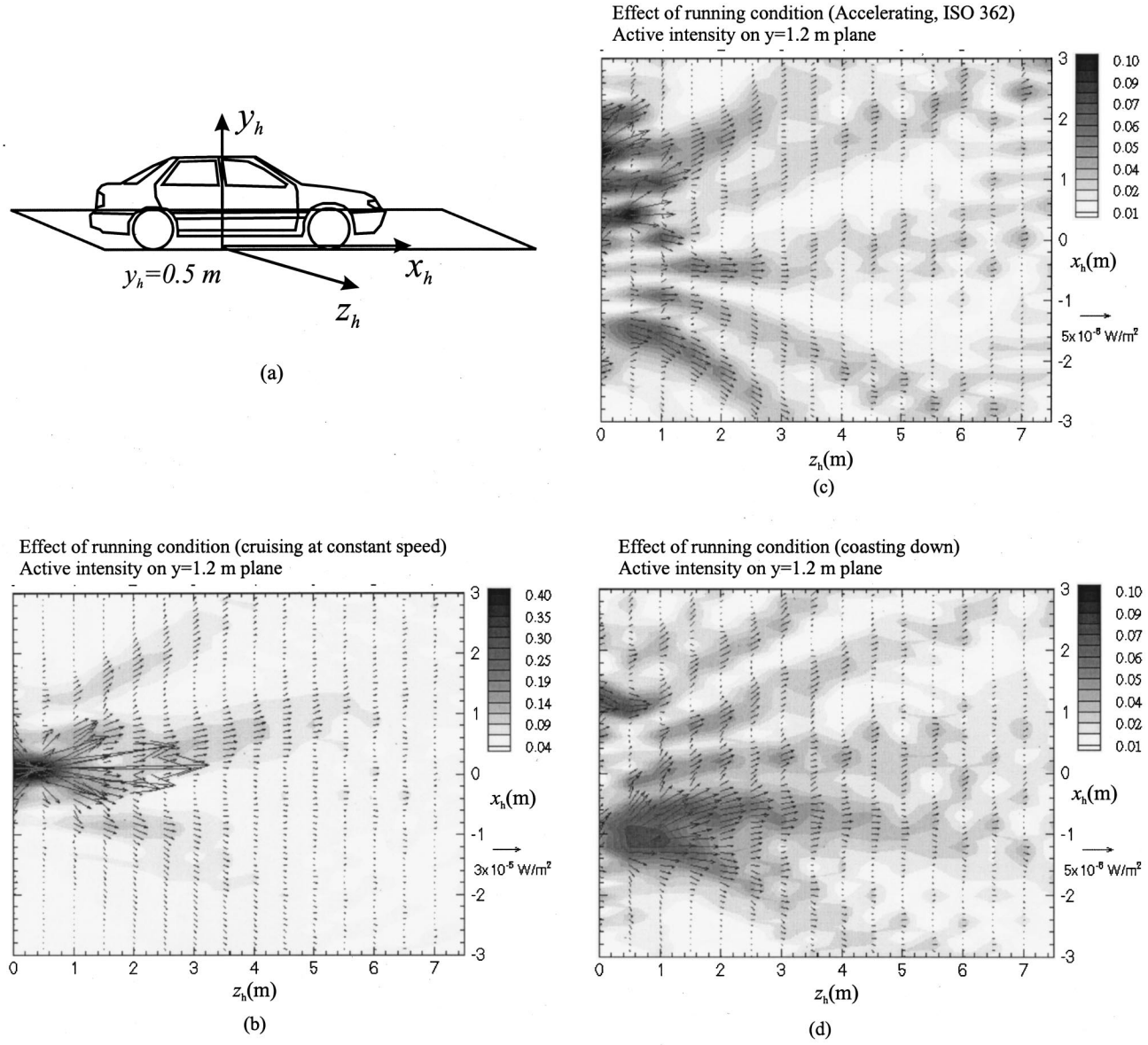


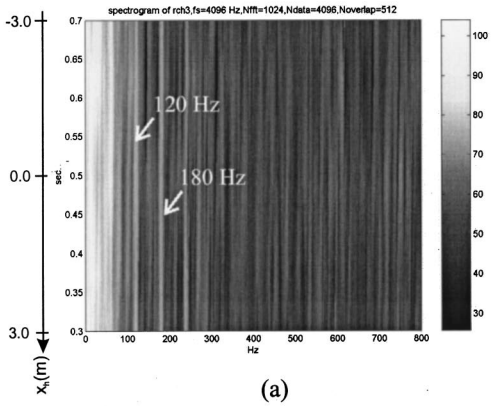
FIG. 16. The propagation of active intensity and sound pressure radiated by the special pattern tire to the far field. (a) The active intensity (x and z direction) and sound pressure are visualized on y=0.5 m plane. (b) Constant velocity condition. (c) Accelerating condition. (d) Coast down condition.

14(a)]. Figure 14(b) shows the direction of propagation of the radiated noise. The radiated sound from the front tire propagates in the forward direction. Figure 15 shows the result when the vehicle coasts down from 55 km/h. The result illustrates that the rear tire is dominant noise source. We can observe that the radiated sound varies significantly according to the driving condition of the vehicle. Figure 16 shows active intensity (x and z directional component) and sound pressure from the source to the pass-by noise measurement location. They are visualized on the $y_h=0.5$ m plane [Fig. 16(a)]. The results demonstrate how the radiated noise propagates in the far field.

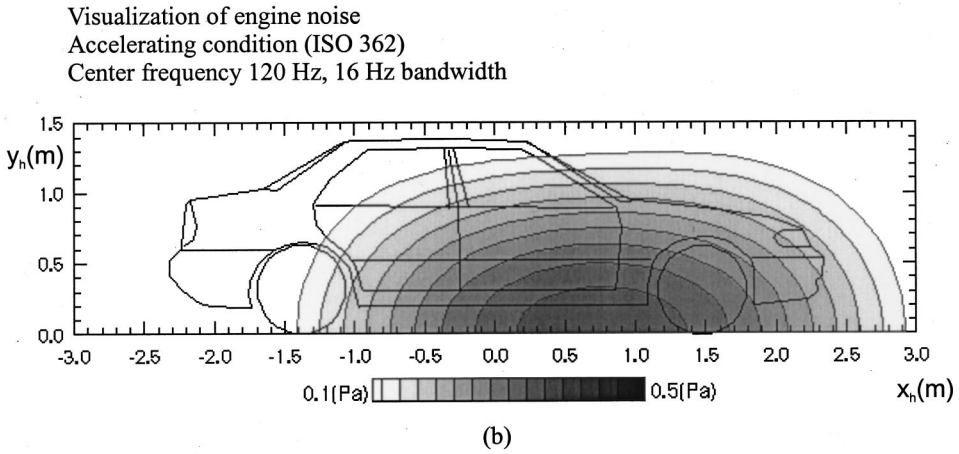
Figures 17 and 18 illustrate the engine noise of a vehicle during pass-by test. The spectrogram of the noise is shown in Fig. 17(a). The frequency change is very small so that the proposed method can be applied. The radiated noise at the second and the third harmonic of firing frequency of engine (60 Hz) were visualized in Figs. 17(b) and (c). Notice that the rear seat is very noisy at the third harmonic of the firing frequency. The three dimensional active intensity distributions on the assumed source plane are visualized in Fig. 18.

IV. CONCLUSIONS

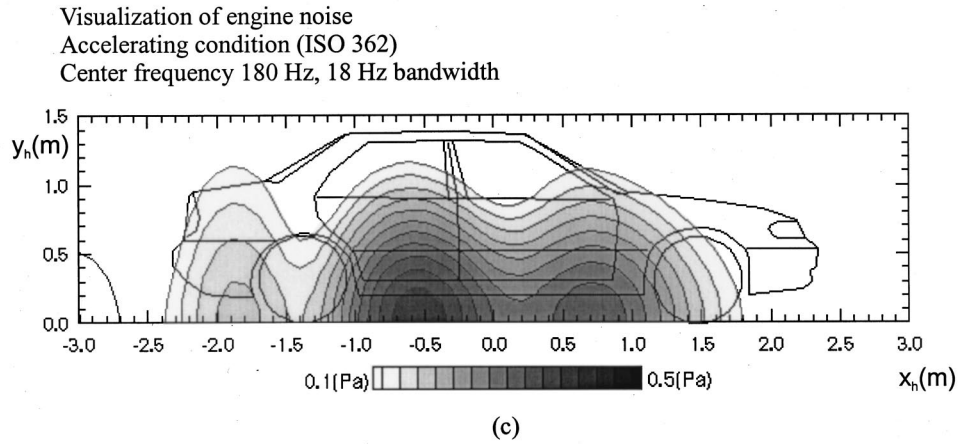
Tire and engine noise during pass-by test can be visualized by means of an improved moving frame acoustic holography (MFAH). We assumed that tire noise during pass-by test is quasistationary. The proposed method was verified by a loudspeaker experiment. The magnitude correction method was also proposed. A special tread pattern tires were made in order to visualize effects of driving condition. The driving condition significantly changes the radiation pattern. Sound intensity maps were also visualized from the source plane to the measurement position of the pass-by noise level (ISO 362).



(a)



(b)



(c)

FIG. 17. Engine noise distribution (sound pressure magnitude) on the assumed source plane (the surface of vehicle, $z_h=0$). (a) Spectrogram of the engine noise. (b) Pressure distribution of the engine noise at 120 Hz. The bandwidth is 16 Hz. (c) Pressure distribution of the engine noise at 180 Hz. The bandwidth is 18 Hz.

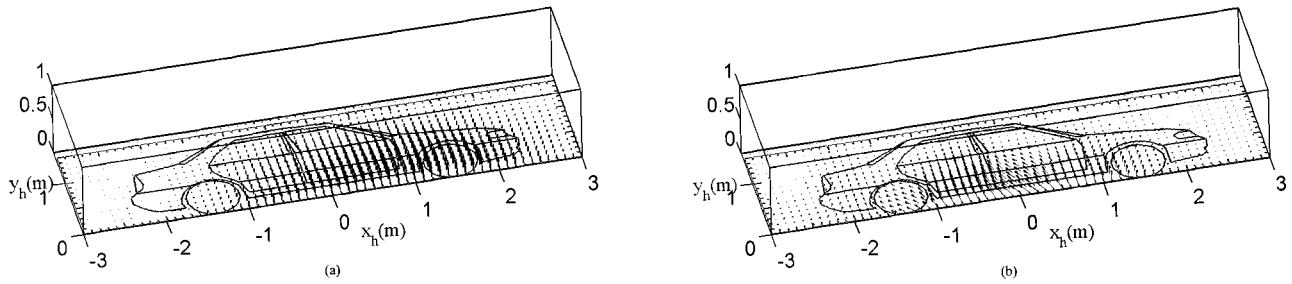


FIG. 18. Engine noise distribution (active intensity) on the assumed source plane (the surface of vehicle, $z_h=0$). (a) Active intensity distribution of the engine noise at 120 Hz. The bandwidth is 16 Hz. (b) Active intensity distribution of the engine noise at 180 Hz. The bandwidth is 18 Hz.

ACKNOWLEDGMENTS

The authors would like to thank Dr. S. J. Lee, H. J. Kim, and S. K. Lee for their support for experiments. This work is supported by Hankook Tire Co., Ltd., the Korea Institute of Science and Technology Evaluation and Planning (KISTEP, "National Research Laboratory" Project), and the Ministry of Education ("Brain Korea 21" Project).

APPENDIX: THE FEASIBILITY OF USING REFERENCE MICROPHONE FOR THE MAGNITUDE CORRECTION

Let us consider N perfectly correlated noise sources that emit a coherent sound field. We denote the spectrum of nth source located at (x_{Sn},y_{Sn},z_{Sn}) by $S_n(f_h)$ $(n=1,2,\ldots,N)$. Since the sources are perfectly correlated the following relation must be satisfied:

$$S_n(f_h)=G_{n1}(f_h)S_1(f_h) \quad (n=2,3,\ldots,N), \tag{A1}$$

where the transfer function $G_{n1}(f_h)$ can be assumed to be constant $(G_{n1}(f_h)=G_{n1}^0)$ for narrow band signal. We denote the transfer function between the nth source signal and the signal at (x_h,y_h,z_H) on the hologram as $H_n(R_n,f_h)$, where $R_n=\sqrt{(x_h-x_{Sn})^2+(y_h-y_{Sn})^2+(z_n-z_{Sn})^2}$ (see Fig. 19). When the source emits a narrow band noise, we can assume that $H_n(f_h)$ is independent of frequency. In this case, we denote $H_n(R_n,f_h)=H_n^0(R_n)$. Then, the pressure spectrum at a point on the hologram can be written as

$$P_{h,k}(x_h,y_h,z_H;k\Delta f_h)=\sum_n H_n(R_n,k\Delta f_h)S_n(k\Delta f_h)$$

$$=\sum_n H_n^0(R_n)S_n(k\Delta f_h), \tag{A2}$$

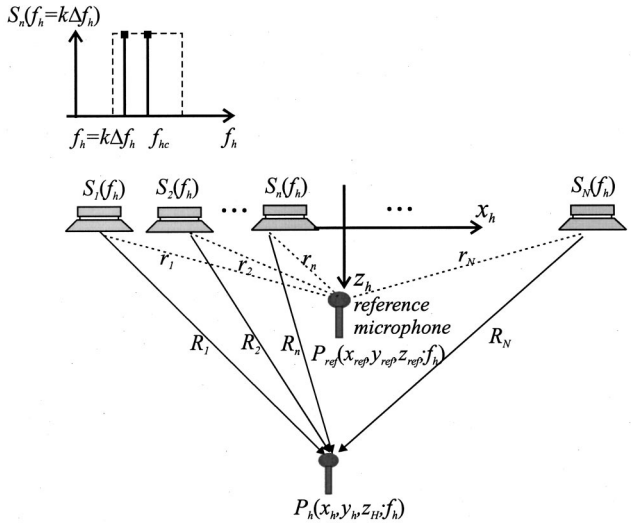


FIG. 19. N perfectly correlated noise sources that emit a coherent sound field.

where $f_k=k\Delta f_h$ (k is integer). At the center frequency ($f_h=f_{hc}$), we rewrite Eq. (A2) as

$$P_{hc}(x_h,y_h,z_H;f_{hc})=\sum_n H_n(R_n,f_{hc})S_n(f_{hc})$$

$$=\sum_n H_n^0(R_n)S_n(f_{hc}), \tag{A3}$$

by using the notation previously used in Sec. II. Then, the correction factor defined by Eq. (17) can be expressed by means of Eqs. (A2) and (A3),

$$\alpha=\sqrt{\sum_k |P_{h,k}(x_h,y_h,z_h;k\Delta f_h)|^2 \Big/ |P_{hc}(x_h,y_h,z_h;f_{hc})|^2}$$

$$=\sqrt{\sum_k \left|\sum_n H_n^0(R_n)G_{n1}^0S_1(k\Delta f_h)\right|^2 \Big/ \left|\sum_n H_n^0(R_n)G_{n1}^0S_1(f_{hc})\right|^2}$$

$$=\sqrt{\left|\sum_n H_n^0(R_n)G_{n1}^0\right|^2 \sum_k |S_1(k\Delta f_h)|^2 \Big/ \left|\sum_n H_n^0(R_n)G_{n1}^0\right|^2 |S_1(f_{hc})|^2}=\sqrt{\sum_k |S_1(k\Delta f_h)|^2/|S_1(f_{hc})|^2}. \tag{A4}$$

This equation can also be obtained from the reference microphone spectrum when the noise sources radiate narrow band noise. We denote the transfer function between the nth source signal, $S_n(f_h)$, and the signal at a reference microphone $(x_{\rm ref},y_{\rm ref},z_{\rm ref})$ as $T_n(r_n,f_h)$, where $r_n=\sqrt{(x_{\rm ref}-x_{Sn})^2+(y_{\rm ref}-y_{Sn})^2+(z_{\rm ref}-z_{Sn})^2}$ (see Fig. 19 again). As previously mentioned, $T_n(r_n,f_h)$ is independent of frequency for narrow band noise, that is, $T_n(r_n,f_h)=T_n^0(r_n)$. The spectrum at the reference microphone can be written as

$$P_{{\rm ref},k}(x_{\rm ref},y_{\rm ref},z_{\rm ref};k\Delta f_n)=\sum_n T_n(r_n,k\Delta f_h)S_n(k\Delta f_h)=\sum_n T_n^0(r_n)S_n(k\Delta f_h). \tag{A5}$$

At the center frequency,

$$P_{{\rm ref}c}(x_{\rm ref},y_{\rm ref},z_{\rm ref};f_{hc})=\sum_n T_n(r_n,f_{hc})S_n(f_{hc})=\sum_n T_n^0(r_n)S_n(f_{hc}). \tag{A6}$$

From Eqs. (A5) and (A6), Eq. (18) can be expressed as

$$\alpha_{\mathrm{ref}} = \sqrt{\sum_k |P_{\mathrm{ref},k}(x_{\mathrm{ref}}, y_{\mathrm{ref}}, z_{\mathrm{ref}}; k\Delta f_h)|^2 / |P_{\mathrm{ref}c}(x_{\mathrm{ref}}, y_{\mathrm{ref}}, z_{\mathrm{ref}}; f_{hc})|^2}$$

$$= \sqrt{\sum_k \left| \sum_n T_n^0(r_n) S_n(k\Delta f_h) \right|^2 \Big/ \left| \sum_n T_n^0(r_n) S_n(f_{hc}) \right|^2}. \tag{A7}$$

If we use $S_n(f_h) = G_{n1}^0 S_1(f_h)$, Eq. (A7) can be reduced to

$$\alpha_{\mathrm{ref}} = \sqrt{\sum_k \left| \sum_n T_n^0(r_n) G_{n1}^0 S_1(k\Delta f_h) \right|^2 \Big/ \left| \sum_n T_n^0(r_n) G_{n1}^0 S_1(f_{hc}) \right|^2}$$

$$= \sqrt{\left| \sum_n T_n^0(r_n) G_{n1}^0 \right|^2 \sum_k |S_1(k\Delta f_h)|^2 \Big/ \left| \sum_n T_n^0(r_n) G_{n1}^0 \right|^2 |S_1(f_{hc})|^2} = \sqrt{\sum_k |S_1 k\Delta f_h)|^2 / |S_1(f_{hc})|^2}. \tag{A8}$$

Therefore, we can prove Eq. (19) by comparing Eq. (A8) with Eq. (A4).

[1] ISO362, ''Acoustics—Measurement of noise emitted by accelerating highway vehicles—Engineering method.''

[2] SAEJ1470, ''Measurement of noise emitted by accelerating highway vehicles.''

[3] W. F. King III and D. Bechert, ''On the sources of wayside noise generated by high-speed trains,'' J. Sound Vib. **66**, 311–332 (1979).

[4] B. Barsikow, W. F. King III, and E. Pfizenmaier, ''Wheel/rail noise generated by a high-speed train investigated with a line array of microphones,'' J. Sound Vib. **118**, 99–122 (1987).

[5] B. Barsikow and W. F. King III, ''On removing the Doppler frequency shift from array measurements of railway noise,'' J. Sound Vib. **102**, 190–196 (1988).

[6] T. Tanaka, I. Sakamoto, T. Murakami, S. Mori, and S. Ogata, ''Measurement of tyre/road noise of a running vehicle by acoustical holography,'' in *Proceedings of Inter Noise 91* (Sydney, Australia, 1991), pp. 1081–1084.

[7] A. Torri, Y. Takano, K. Terada, M. Sebata, A. Iida, K. Horihata, and H. Fujita, ''Shinkansens's sound source measurements using microphone arrays,'' in *Proceedings of Inter Noise 92* (Toronto, Canada, 1992), pp. 1171–1174.

[8] Y. Takano, K. Terada, F. Aizawa, A. Iida, and H. Fujita, ''Development of a 2-dimensional microphone array measurement system for noise sources of fast moving vehicles,'' in *Proceedings of Inter Noise 92* (Toronto, Canada, 1992), pp. 1175–1178.

[9] S. Brühl and K.-P. Schmitz, ''Noise source localization on highspeed trains using different array types,'' in *Proceedings of Inter Noise 93* (Leuven, Belgium, 1993), pp. 1311–1314.

[10] J. F. Harmet, M. A. Pallas, and K.-P. Schmitz, ''Deufrako-1: Microphone array techniques used to locate acoustic sources on ICE, TGV-V and TRANSRAPID 07,'' in *Proceedings of Inter Noise 94* (Yokohama, Japan, 1994), p. 187–192.

[11] G. Hölzl, P. Fodiman, K.-P. Schmitz, M. A. Pallas, and B. Barsikow, ''Deufrako-2: Localized sound source on the high-speed vehicles on ICE, TGV-V and TRANSRAPID 07,'' in *Proceedings of Inter Noise 94* (Yokohama, Japan, 1994), pp. 193–198.

[12] I. Sakamoto and T. Tanaka, ''Application of acoustic holography to measurement of noise on an operating vehicle,'' SAE paper No. 930199 (1993).

[13] S.-M. Kim, H.-S. Kwon, S.-H. Park, and Y.-H. Kim, ''Experimental comparisons of sound visualization methods for moving noise sources,'' in *Proceedings of Inter Noise 96* (Liverpool, UK, 1996), pp. 377–380.

[14] H. Nakagawa, T. Tsuru, T. Tanaka, and I. Sakamoto, ''Detection and visualization of moving sound source through acoustic holography,'' in *Proceedings of ASVA 97* (Tokyo, Japan, 1997), pp. 591–594.

[15] F. Poisson, J. C. Valiere, and P. Herzog, ''High speed sound sources localization using bilinear time-frequency transformation,'' Appl. Acoust. **53**, 1–13 (1998).

[16] W. M. To and S. M. Yung, ''Power estimation of sound sources on low-speed electric trains using a deconvolution approach,'' J. Acoust. Soc. Am. **105**, 2350–2358 (1999).

[17] J. D. Maynard, E. G. Williams, and Y. Lee, ''Near field acoustic holography: I. Theory of generalized holography and development of NAH,'' J. Acoust. Soc. Am. **78**, 1395–1413 (1985).

[18] P. Rasmussen, S. Gade, and G. Ginn, ''Tire noise measurement on a moving vehicle,'' Fourth International Congress on Sound and Vibration, 1996, pp. 599–602.

[19] H.-S. Kwon and Y.-H. Kim, ''Moving frame technique for planar acoustic holography,'' J. Acoust. Soc. Am. **103**, 1734–1742 (1998).

[20] S.-H. Park and Y.-H. Kim, ''An improved moving frame acoustic holography for coherent bandlimited noise,'' J. Acoust. Soc. Am. **104**, 3179–3189 (1998).

[21] S.-H. Park and Y.-H. Kim, ''Effects of the speed of moving noise sources on the sound visualization by means of moving frame acoustic holography (MFAH),'' J. Acoust. Soc. Am. **108**, 2719–2728 (2000).

[22] E.-U. Saemann and J. Hald, ''Transient tyre noise measurement using time domain acoustic holography,'' SAE paper No. 972050 (1997).

[23] J. Hald, STSF-a unique technique for scan-based Near-field Acoustical Holography without restrictions on coherence, Brüel & Kjaer Technical Review No. 1, 1989.

Pulse response of a nonlinear layer

Claes M. Hedberg[a)]
Blekinge Institute of Technology, 371 79 Karlskrona, Sweden

Oleg V. Rudenko
Moscow State University, Moscow 119899, Russia and Blekinge Institute of Technology, 371 79 Karlskrona, Sweden

(Received 19 December 2000; accepted for publication 15 July 2001)

A simple analytical theory is developed for the description of the non-steady-state response of a thin nonlinear layer, which differs markedly in its linear properties from the surrounding medium. Such a layer can model the behavior of real inhomogeneities like a cloud of gas bubbles in a liquid, a crack or split plane in a solid, or the contact between two slightly tightened rough surfaces. Both weakly nonlinear pulse and harmonic responses are calculated and the general properties of the spectral and temporal structures of the scattered field are discussed. Exact strongly nonlinear solutions are derived for a special type of stress–strain relationship corresponding to the behavior of real condensed media under strong loads. Profiles and spectra shown conform with experimental results. The pulse response on the short δ-pulse shaped incident wave is calculated for arbitrary nonlinear properties of the layer. The possibilities to apply the sets of data on measured characteristics of pulse response in the solution of inverse problems are briefly discussed. [DOI: 10.1121/1.1402117]

PACS numbers: 43.25.Ba, 43.25.Jh [MFH]

I. INTRODUCTION

It is well known that a huge nonlinear acoustic response is obtained from liquids containing air bubbles.[1–3] The response indicates a magnitude of nonlinearity ϵ for the gas–liquid mixture of up to 10^3,[1] which far exceeds the nonlinearity of homogeneous fluids. For example, water has ($\epsilon \approx 3.5$) and air has ($\epsilon = 1.2$). The high sensitivity of nonlinear testing has application in the detection of microbubbles, early in the development of boiling or in the detection of fermentation processes. The medical agent ''Albunex'' is used for nonlinear ultrasonic imaging of the blood flow;[4] similarly, nitrogen bubbles can be detected in the blood of divers.

The huge increase in nonlinear response of solids containing defects in their structure (cracks, loosely packed grains, pores filled in part by gas and liquid, dislocations, residual stresses, etc.) has found widespread use in the testing of composites, buildings, structural materials, industrial products, and rocks (see, for example, Refs. 5–9).

In the last few years new interesting aspects have sparked the interest in nonlinear nondestructive testing (NNDT).[10,11] A review of experimental results on material characterization including nonlinear reflection methods, is given in Ref. 12.

Needless to say, NNDT calls for the solution of inverse problem, i.e., the reconstruction of shapes, sizes, or other physical properties of inhomogeneities using measurements of the scattered field. In general, the inverse problem of NNDT can be related to nonlinear diffraction tomography.[13] However, supplementary *a priori* information is necessary for efficient solution of the inverse problem, which can result from the solution of direct problems on nonlinear transmission, reflection, or scattering. Unfortunately the set of available direct data leaves something to be desired.

As recent attention has focused on NNDT, extension of the known linear solutions to the corresponding nonlinear problems is a natural path to follow. The problem of normal incidence of a plane wave on a layer is one of the simplest and most important problems in the acoustics of layered media.[14] It attracts considerable interest for two reasons. First, it is rather simple, and one can look forward to obtaining the solution, which must be in analytical form. Second, the layer can serve as a model of a real nonlinear inclusion such as a cloud of bubbles in a liquid, or a region inside a solid with a high content of defects, or a resonant cavity in concrete, or a geological structure. However, the nonlinear problem is complicated, even if the nonlinear phenomena are calculated by the successive approximation method. The main difficulty is in the scattered field, which cannot be considered monochromatic. Even if the incident wave contains a single harmonic, higher ones will be generated in the layer. A more interesting problem is multifrequency incidence where one of the combination frequencies excites a resonant vibration of the layer.[15] Finally, it is easy to reach a well-defined nonlinear response with a high-intensity pulse signal acting on the layer.

II. GENERAL FORMULATION OF THE PROBLEM

The statement of the problem is illustrated in Fig. 1. A plane layer with density ρ_0 and sound velocity c_0 is placed between $x=0$ and $x=h$. This layer is surrounded by another medium with density ρ_1 and sound velocity c_1. The relationship between the acoustical properties of both media can be arbitrary. Significant attention will be focused in the next sections on a soft and thin layer

a)Electronic mail: claes.hedberg@ima.hk-r.se

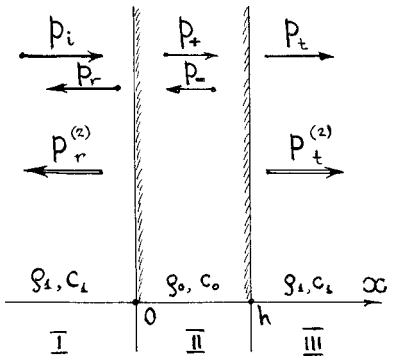


FIG. 1. Layer ($0<x<h$), surrounding medium, and waves taken into account at the formulation of the problem.

$$\zeta=\frac{c_0\rho_0}{c_1\rho_1}\ll 1, \quad \frac{\omega}{c_0}h\ll 1. \tag{1}$$

Also, the nonlinear properties of the layer will be taken into consideration.

It is presumed that the geometric nonlinearity (see, for example, Ref. 15, pp. 39–40) initiated from the structure of the equations of motion and continuity is small in comparison with the physical nonlinearity caused by the finite deformation of the layer. Therefore, these equations can be reduced to a linear connection between the variations of pressure p and density ρ

$$\nabla^2 p-\frac{\partial^2\rho}{\partial t^2}=0. \tag{2}$$

But, the ''equation of state''

$$\rho=\rho(p), \tag{3}$$

must be nonlinear. Equation (3) can be algebraic, for example a power series in p, or a more complicated function which is necessary in the case of strong nonlinearity. If the medium has a ''memory,'' the Taylor series must be replaced by the corresponding functional expansion of Volterra–Frechet type,[15] pp. 80–81. Other models, for example accounting for the hysteretic properties, can be considered by a simplified approach in (2), (3).

The acoustic field in the linear medium surrounding the layer can be written as the sum of two waves traveling in opposite directions. In the region I, $-\infty<x<0$, the field consists of incident p_i and reflected p_r waves (see Fig. 1)

$$\begin{aligned} &p^{\mathrm{I}}=p_i(t-x/c_1)+p_r(t+x/c_1),\\ &\rho_1 c_1 u^{\mathrm{I}}=p_i(t-x/c_1)-p_r(t+x/c_1). \end{aligned} \tag{4}$$

The field in the region III, $x>h$, is formed by the transmitted wave

$$\begin{aligned} &p^{\mathrm{III}}=p_t(t-(x-h)/c_1),\\ &\rho_1 c_1 u^{\mathrm{III}}=p_t(t-(x-h)/c_1). \end{aligned} \tag{5}$$

When the nonlinear properties of the layer are significant, both the reflected p_r and transmitted p_t waves will contain nonlinear components, $p_r^{(2)}$ and $p_t^{(2)}$, respectively (see Fig. 1).

The acoustic field inside the layer (region II in Fig. 1) is described by the nonlinear equations (2) and (3). In the linear approximation it consists of right-going p_+ and left-going p_- waves

$$\begin{aligned} &p^{\mathrm{II}}=p_+(t-x/c_0)+p_-(t+x/c_0),\\ &\rho_0 c_0 u^{\mathrm{II}}=p_+(t-x/c_0)-p_-(t+x/c_0). \end{aligned} \tag{6}$$

The brief solution for the linear problem is given below to illustrate the mathematical technique and to derive the necessary formulas for the pulse response which will be useful for the nonlinear study.

Because most of this work deals with nonlinear pulse waves, we will avoid Fourier analysis also when solving the linear problem.

III. LINEAR PULSE RESPONSE OF A LAYER

To calculate the temporal profiles of both reflected and transmitted waves, the fields in the three regions (4)–(6) must be sewed together. The pressures and velocities must be equal at both boundaries, $x=0$ and $x=h$. These boundary conditions lead to the functional equations

$$p_i(t)+p_r(t)=p_+(t)+p_-(t), \tag{7}$$

$$\zeta[p_i(t)-p_r(t)]=p_+(t)-p_-(t), \tag{8}$$

$$\rho_+(t-h/c_0)+p_-(t+h/c_0)=p_t(t), \tag{9}$$

$$p_+(t-h/c_0)-p_-(t+h/c_0)=\zeta p_t(t), \tag{10}$$

used in the calculation of the unknown profiles p_r, p_t, p_+, and p_-. The incident pulse $p_i(t)$ is assumed to be known.

The solution of the system (7)–(10) is easily found

$$p_+(t)=\zeta(1+\zeta)\hat{G}\cdot \mathrm{e}^{\hat{D}h/c_0}\cdot p_i(t), \tag{11}$$

$$p_-(t)=\zeta(1-\zeta)\hat{G}\cdot \mathrm{e}^{-\hat{D}h/c_0}\cdot p_i(t), \tag{12}$$

$$p_t(t)=2\zeta\hat{G}\cdot p_i(t), \tag{13}$$

$$p_r(t)=-(1-\zeta^2)\hat{G}\cdot \sinh(\hat{D}h/c_0)\cdot p_i(t). \tag{14}$$

A symbolic operator-form representation is used, where

$$\hat{G}=\left[(1+\zeta^2)\sinh\left(\frac{h}{c_0}\hat{D}\right)+2\zeta\cosh\left(\frac{h}{c_0}\hat{D}\right)\right]^{-1}, \quad \hat{D}=\frac{\partial}{\partial t}, \tag{15}$$

and $\mathrm{e}^{\tau\hat{D}}$ is a standard shift operator

$$\mathrm{e}^{\tau\hat{D}}p(t)\doteq p(t+\tau). \tag{16}$$

The acoustic pressure field inside the layer (4) equals

$$p^{\mathrm{II}}(t)=2\zeta\left\{\cosh\left[\left(\frac{h-x}{c_0}\right)\hat{D}\right]+\zeta\sinh\left[\left(\frac{h-x}{c_0}\right)\hat{D}\right]\right\}\cdot\hat{G}\cdot p_i(t). \tag{17}$$

For a single harmonic spectrum, $p_i=A_i\exp(-i\omega t)$, the operators rearrange into

$$\hat{D}\to -i\omega, \quad \cosh\left(\frac{h}{c_0}\hat{D}\right)\to\cos(\kappa_0 h),$$
$$\sinh\left(\frac{h}{c_0}\hat{D}\right)\to -i\sin(\kappa_0 h), \tag{18}$$

and the solution (17) takes the common form[14]

$$p^{\mathrm{II}}=2\zeta\frac{\zeta\sin[\kappa_0(h-x)]+i\cos[\kappa_0(h-x)]}{(1+\zeta^2)\sin(\kappa_0 h)+i2\zeta\cos(\kappa_0 h)}A_i e^{-i\omega t}. \tag{19}$$

At significant difference in impedances, $\zeta\ll 1$ or $\zeta\gg 1$, the resonant properties of the layer are clearly defined at

$$\kappa_0 h=\pi n, \quad h=n\frac{\lambda}{2} \quad (n=1,2,3,...). \tag{20}$$

The response of the layer is larger for the frequency components of the incident spectrum satisfying the condition (20).

A resonant behavior of a layer is favorable in the generation of new nonlinear frequencies. However, this problem calls for an extended and special investigation which is beyond the scope of the present work. On the other hand, NNDT often deals with nonresonant inclusions, which are small in comparison with the wavelength. To solve this important group of problems, a special limiting case of Eqs. (11)–(15) must be considered. Namely, both the ratio of impedances ζ (1) and wave thickness of layer $\omega h/c_0$ must simultaneously be small parameters. None is infinitesimally small, because the ratio $c_0\zeta/\omega h$ can be arbitrary in its magnitude. For such a thin inhomogeneity the solution (11)–(15) can be simplified with account for the inequalities (1)

$$\hat{G}^{-1}\approx\frac{h}{c_0}\hat{D}+2\zeta, \tag{21}$$

$$\left(\frac{d}{dt}+\frac{2c_0\zeta}{h}\right)\begin{pmatrix} p_{\pm} \\ p_t \\ p_r \end{pmatrix}=\begin{pmatrix} c_0\zeta/h \\ 2c_0\zeta/h \\ -d/dt \end{pmatrix}p_i(t). \tag{22}$$

The most important properties of the response of a thin layer are the following ones.

The profile of the transmitted pulse equals

$$p_t(t)=\frac{2c_0\zeta}{h}\int_{-\infty}^{t}p_i(t')e^{-2c_0\zeta(t-t')/h}dt'. \tag{23}$$

It follows from (23), that $h/2c_0\zeta$ plays the role of ''relaxation time'' of a layer. The combination $\omega h/2c_0\zeta=a^{-1}$ used later can be considered as a normalized relaxation time. The reflected pulse is calculated by the formula

$$p_r(t)=-\frac{h}{2c_0\zeta}\frac{dp_t}{dt}=p_t-p_i. \tag{24}$$

One can see that the solutions (23) and (24) satisfy the energy conservation law for pulse response, namely

$$\int_{-\infty}^{\infty}p_i^2(t)dt=\int_{-\infty}^{\infty}[p_r^2(t)+p_t^2(t)]dt. \tag{25}$$

The fulfillment of the condition

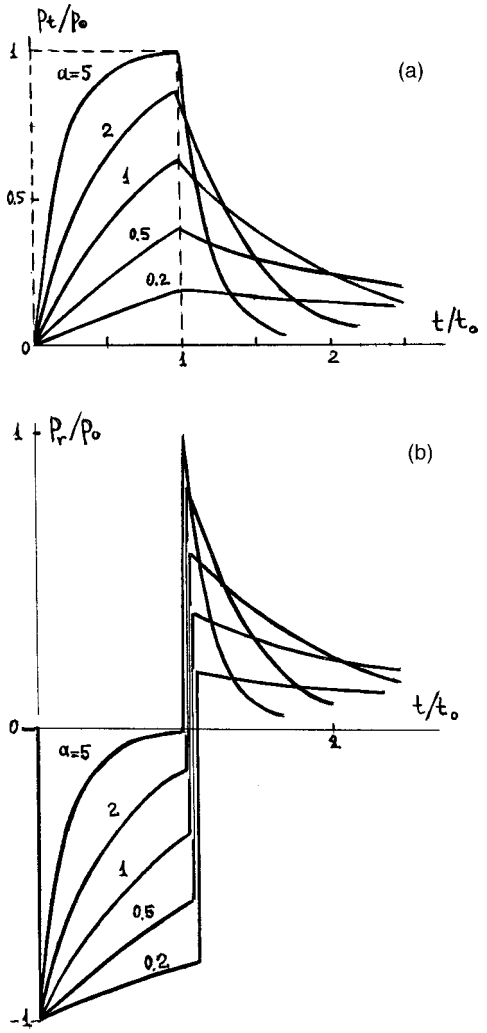


FIG. 2. Profiles of transmitted (a) and reflected (b) pulses at different values of the parameter $a=2c_0t_0\zeta/h$. The incident pulse of rectangular form is shown by the dashed line.

$$\int_{-\infty}^{\infty}p_r(t)p_t^n(t)dt=0, \tag{26}$$

means, in particular (at $n=1$), that the forms of reflected and transmitted pulses are orthogonal to each other.

The profiles of right- and left-going waves in the layer are equal, $p_+(t)=p_-(t)$. The inner acoustic field

$$p^{\mathrm{II}}(t)=p_t(t)=p_i(t)+p_r(t), \tag{27}$$

has the same time wave form as the transmitted pulse.

The profiles of the transmitted p_t and reflected p_r pulses are shown in Figs. 2(a) and (b) for different values of the parameter $a=2c_0t_0\zeta/h$ equal to 0.2, 0.5, 1, 2, and 5. Here, t_0 is the duration of the incident pulse of rectangular form shown by the dashed line in Fig. 2(a). It is significant that both the leading and the tail fronts of the monopolar transmitted pulse are smoothed out. On the other hand, the leading fronts of both negative and positive sections of the bipolar reflected pulse are as steep as the front of the incident pulse.

It is convenient to keep in mind some numerical estimations. If a gaseous layer (air) is surrounded by a denser medium (water), $\zeta\approx 3\cdot 10^{-4}$, the parameter $a=1$ corresponds to the ratio $h/t_0=20$ cm/s. For example, the width of a layer

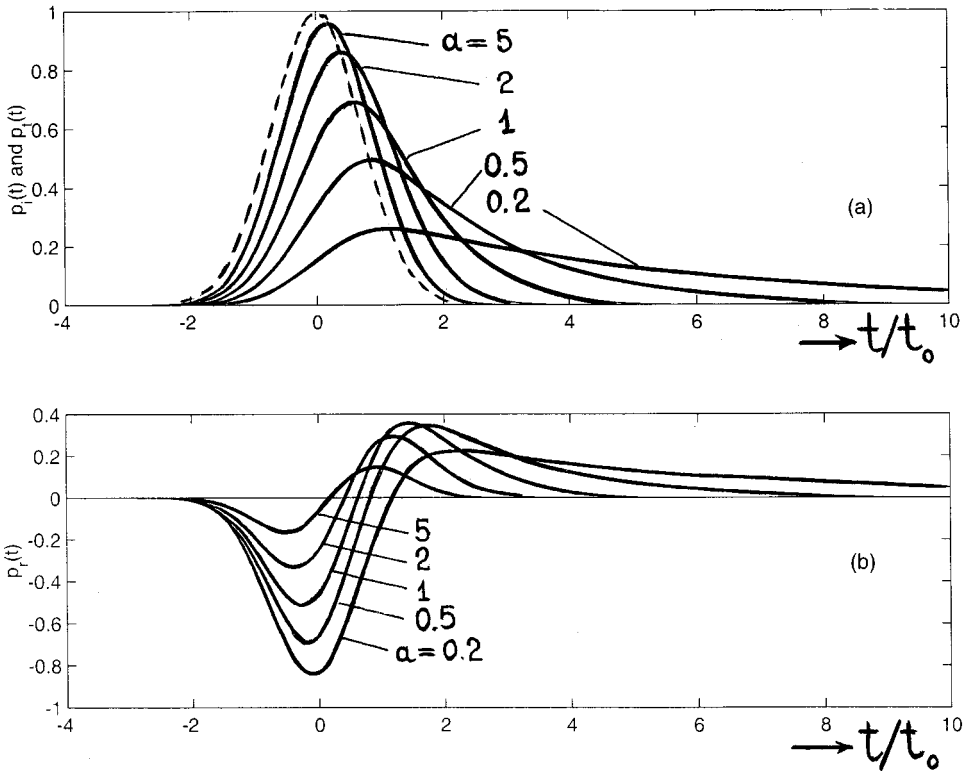


FIG. 3. Response to Gaussian pulse (dashed line): forms of transmitted (a) and reflected (b) pulses.

$h \sim 0.01$ cm corresponds to a duration of the probing pulse in the order of 0.5 ms.

In Fig. 3 the reflected and transmitted pulses are shown for the same parameters a, as in Fig. 2, but for a Gaussian incident wave

$$p_i = p_0 \exp(-t^2/t_0^2); \tag{28}$$

see the dashed line in Fig. 3(a). The main peculiarities of the pulse response exhibited in Fig. 2 can also be seen in Fig. 3.

It should be emphasized that the simple analytical results (22)–(24) shown in Figs. 2 and 3 are derived with account for both inequalities (1). The small parameters (1) are presented in all solutions in the form of a ratio, equal to $a = 2c_0 t_0 \zeta/h$ for pulse signals, and $a = 2c_0\zeta/h\omega$ for harmonic signals. This ratio of two small numbers can be arbitrary in magnitude.

If the ratio of impedances ζ (1) is very small and $a \to 0$, it follows from the solution (23), (24) that $p_t(t) \to 0$, and $p_r(t) \to -p_i(t)$; this limiting case corresponds to reflection from a free boundary. In the opposite situation, if the thickness h of the layer is very small and $a \to \infty$, it follows from (23), (24) that $p_t(t) \to p_i(t)$, and $p_r(t) \to 0$; this second limit corresponds to complete transmission of the incident wave through the layer.

When assuming a temporal dependence of p_i, p_r, and p_t of the form $\exp(-i\omega t)$, the transmission and reflection coefficients are calculated to be

$$K_t(\omega) = i\frac{2c_0\zeta/\omega h}{1 + i2c_0\zeta/\omega h}, \quad K_r(\omega) = -\frac{1}{1 + i2c_0\zeta/\omega h}. \tag{29}$$

One can see that the high frequencies in the pulse spectrum are reflected quite well, because $K_r \to -1$ at $\omega \to \infty$. So, the steep leading front of the rectangular incident pulse reflects completely from the layer, as is shown in Fig. 2(b).

The last linear example is devoted to the response of a layer to a sinusoidal incident pulse of finite duration, $0 < t < t_0$. The form of the transmitted wave is

$$\begin{aligned} \frac{p_t}{p_0} &= \frac{a^2}{1+a^2}\sin\omega t + \frac{a}{1+a^2}(e^{-a\omega t} - \cos\omega t), \quad 0 < t < t_0, \\ \frac{p_t}{p_0} &= \frac{a^2}{1+a^2}e^{-a\omega(t-t_0)} \cdot \sin\omega t_0 + \frac{a}{1+a^2}e^{-a\omega t} \\ &\quad \times (1 - e^{a\omega t_0}\cos\omega t_0), \quad t > t_0. \end{aligned} \tag{30}$$

As distinct from monopolar incident pulses, the notation for a is here $a = 2c_0\zeta/\omega h$. When $t_0 \to \infty$ and $a\omega t \gg 1$, the result is a steady-state harmonic response

$$\begin{pmatrix} p_t \\ p_r \end{pmatrix} = -\frac{p_0}{\sqrt{\omega^2 + (2c_0\zeta/h)^2}} \begin{pmatrix} 2c_0\zeta/h \\ \omega \end{pmatrix} \begin{pmatrix} \cos \\ \sin \end{pmatrix} \times \left(\omega t + \arctan\frac{2c_0\zeta}{\omega h} \right). \tag{31}$$

The amplitude of the transmitted wave decreases with increase in frequency ω (from p_0 to 0). The amplitude of the reflected wave increases monotonically and tends to p_0 when $\omega \to \infty$.

IV. WEAKLY NONLINEAR THIN LAYER

If the width of the layer h is small in comparison with the wavelength, the inner pressure will be equal to the outer one; in other words: the connection (27) is true also for the nonlinear deformation of the layer

$$p^{\mathrm{I}}=p_i(t)+p_r(t)=p^{\mathrm{II}}=p^{\mathrm{III}}=p_t(t). \tag{32}$$

The boundary conditions (8) and (10) for the particle velocity will be expressed as

$$\frac{1}{\rho_1 c_1}[p_i(t)-p_r(t)]=u(x=0,t), \quad \frac{1}{\rho_1 c_1}p_t(t)=u(x=h,t). \tag{33}$$

It follows from (32) and (33) that the difference of velocities equals

$$u(h,t)-u(0,t)=\frac{2}{\rho_1 c_1}(p^{\mathrm{II}}-p_i). \tag{34}$$

On the other hand, the nonlinear dependence $u(p)$ can be written by means of the linear equation of discontinuity, used at the derivation of (2), and the nonlinear ''equation of state'' (3)

$$\operatorname{div}\mathbf{u}=-\frac{1}{\rho_0}\frac{\partial}{\partial t}\rho(p). \tag{35}$$

For a thin layer (35) takes the form

$$u(h,t)-u(0,t)\approx-\frac{h}{\rho_0}\frac{d}{dt}\rho(p^{\mathrm{II}}). \tag{36}$$

We hereafter symbolize the inner field $p^{\mathrm{II}}\equiv p$. As (34) is equal to (36) we derive the nonlinear equation for the inner field

$$\frac{d}{dt}f(p)+\frac{2c_0\zeta}{h}p=\frac{2c_0\zeta}{h}p_i(t), \quad \rho(p^{\mathrm{II}})\equiv\frac{1}{c_0^2}f(p). \tag{37}$$

The new equation (37) is a generalization of (22) when taking into account the nonlinear properties of the layer. If the inner field $p(t)$ is known, the reflected and transmitted pulses are known as well

$$p_r(t)=p(t)-p_i(t), \quad p_t(t)=p(t). \tag{38}$$

For weak nonlinearity, the function f can be modeled by a series expansion

$$f(p)\approx p-\frac{\epsilon}{c_0^2\rho_0}p^2+\frac{\chi}{(c_0^2\rho_0)^2}p^3+\cdots, \tag{39}$$

where ϵ and χ are the coefficients of quadratic and cubic nonlinearities, respectively. In this case, (37) can be solved by the successive approximation method: $p=p^{(1)}+p^{(2)}+\cdots$. Taking into account only the quadratic term in (39), we obtain a pair of equations for $p^{(1)}$ and $p^{(2)}$

$$\frac{dp^{(1)}}{dt}+\frac{2c_0\zeta}{h}p^{(1)}=\frac{2c_0\zeta}{h}p_i, \tag{40}$$

$$\frac{dp^{(2)}}{dt}+\frac{2c_0\zeta}{h}p^{(2)}=\frac{\epsilon}{c_0^2\rho_0}\frac{d}{dt}p^{(1)^2}. \tag{41}$$

Equation (40) coincides in the first approximation with (22) for the linear transmitted wave p_t. The second-order response is described by an analogous equation, but the ''driving force'' in the right-hand side of (41) is determined by the 1st approximation solution $p^{(1)}$, instead of the incident pulse p_i in Eq. (40).

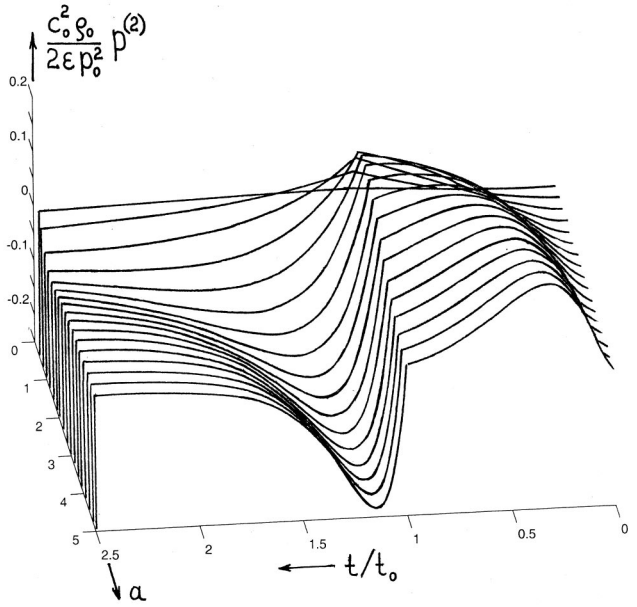


FIG. 4. Weak nonlinear response to rectangular incident pulse. The linear part of response is shown in Fig. 2.

It follows from (41) that the nonlinear pulse is described by the function $p^{(2)}(t)$ whose integral area equals zero. If the monopolar incident pulse has rectangular form, one can calculate the nonlinear response using the solution for $p_t(t)$ [see Fig. 2(a)] as the first approximation

$$p^{(2)}=\frac{2\epsilon p_0^2}{c_0^2\rho_0}e^{-a\tau}\times\begin{cases}(e^{-a\tau}-1+a\tau), & 0<\tau<1,\\ (e^{-a}-1+a)-(e^{a}-1)^2(e^{-a}-e^{-a\tau}), & \tau>1.\end{cases} \tag{42}$$

Here, the parameter $\tau=t/t_0$ and $a=2c_0t_0\zeta/h$. The profiles of the bipolar nonlinear pulse (42) are shown in Fig. 4 for different values of a.

Using the solution (30) as the first approximation, we derive the steady-state nonlinear response to a sinusoidal pulse

$$p^{(2)}=\frac{\epsilon p_0^2}{c_0^2\rho_0}\frac{\omega h/2c_0\zeta}{1+(\omega h/2c_0\zeta)^2}\frac{1}{\sqrt{1+(\omega h/c_0\zeta)^2}}\cdot\cos\left[2\omega t+2\arctan\left(\frac{2c_0\zeta}{\omega h}\right)+\arctan\left(\frac{c_0\zeta}{\omega h}\right)\right]. \tag{43}$$

It is important at this point to evaluate the amplitude of the nonlinear response (43). It reaches its maximum value at

$$\frac{\omega h}{2c_0\zeta}\approx 0.45, \quad p^{(2)}_{\max}\approx 0.3\frac{\epsilon p_0^2}{c_0^2\rho_0}. \tag{44}$$

This is a nonresonant and smooth maximum caused by competition between nonlinear growth of $p^{(2)}$ and the inner field which decays more rapidly with increase in frequency. In the absence of an air layer, a corresponding thin layer of another homogeneous medium can generate a weak second harmonic with the amplitude[15]

$$p_h = \frac{\epsilon_1}{2c_1^3\rho_1}\omega p_0^2 h. \tag{45}$$

This formula can be easily derived from the Bessel–Fubini solution (see, for example, Refs. 1 and 15, p. 31), or from Riemann solution at distances h small in comparison with shock formation distance. The ratio of amplitudes (44) and (45) can reach its maximum possible value

$$\frac{p_{\max}^{(2)}}{p_h} \approx \frac{2}{3}\frac{\epsilon}{\epsilon_1}\left(\frac{c_1^2\rho_1}{c_0^2\rho_0}\right)^2, \tag{46}$$

at $\kappa_0 h \approx 0.9\zeta$.

For an air layer, surrounded by water, the ratio (46) is evaluated as 10^7. The huge increase in the nonlinear properties of a medium containing highly compressible inclusions is impressive, but caution must be exercised in the interpretation of such estimates. For example, the ratio of maximum nonlinear response (44) to an incident amplitude p_0 can be in the order of 1 at pressure $p_0 \sim 20$ atm. Such pressure corresponds in water to a sound intensity of about 10^2 W/cm^2, which is difficult to reach in normal high-power ultrasound experiments. However, there is no problem to form an acoustic pulse with a peak pressure on the order of several hundreds of atmospheres.[16]

A plane cloud of floating bubbles is often used in experiments to create a highly nonlinear region in water. For this purpose, electrolytic decomposition of water molecules into hydrogen and oxygen can be utilized, as well as pumping of compressed air through orifices in a pipeline lying on the bottom.[1–3] Let the width h of the layer be less than the wavelength, and the volume concentration v of air be small. Then, the density of the bubbly liquid is about the density of pure water. It follows from (34) and (36) that

$$\frac{d}{dt}\rho(p) + \frac{2}{c_1 h}p = \frac{2}{c_1 h}p_i(t), \tag{47}$$

where the variation of density caused by the incident pulse p_i is connected only with the compressibility of air

$$\rho = -\rho_1 n V. \tag{48}$$

Here, n is the number of bubbles per unit volume, and V is the variation of the volume of a single bubble described by the equation (see, for example, Ref. 15, p. 160)

$$\frac{d^2V}{dt^2} + \omega_0^2(V - gV^2) = -\nu p. \tag{49}$$

The following notations are used here:

$$\omega_0^2 = \frac{3c_0^2}{R_0^2}, \quad g = \frac{3\epsilon}{4\pi R_0^3}, \quad \nu = \frac{4\pi R_0^2}{\rho_0}, \tag{50}$$

where R_0 is the equilibrium radius of the air bubble. Combining Eqs. (47)–(49), one can derive the equation governing the response of a bubble inside the layer for an incident pulse

$$\frac{d^2V}{dt^2} + \frac{1}{2}\nu\rho_1 c_1 n h \cdot \frac{dV}{dt} + \omega_0^2(V - gV^2) = -\nu p_i. \tag{51}$$

The second term on the left-hand side of (51) is connected with energy lost in the excitation of reflected and transmitted waves. The Q-factor for the oscillatory system (51)

$$Q = \frac{2\omega_0}{\nu\rho_1 c_1 n h} = \frac{2}{\sqrt{3}}\frac{c_0\rho_0}{c_1\rho_1}\frac{R_0}{hv}, \quad v = n\cdot\frac{4}{3}\pi R_0^3, \tag{52}$$

is small in many instances, and then the second derivative can be eliminated. In this case the system (47)–(49) can be reduced to the equation

$$\frac{d}{dt}\left(p - \frac{\epsilon}{c_0^2\rho_0}p^2\right) + \frac{2c_0\zeta}{hv}p = \frac{2c_0\zeta}{hv}p_i(t), \tag{53}$$

which differs in notations only from (37), accounting only for quadratic nonlinearity. The ratio of nonlinear responses (44) and (45) is described by an equation similar to (46), but its right-hand side must be multiplied by the volume concentration v of the gas. Consequently, the evaluation (46) for $v = 10^{-4}$ gives an increase in nonlinearity of 10^3 times (instead of 10^7 indicated above). This value corresponds to measurements.[1]

Considering the more exotic case when $Q > 1$, which means that a cloud of bubbles can execute free quasiperiodic oscillations, the power expansion $\rho(p)$ (39) must be replaced by a functional series of Volterra–Frechet type

$$\begin{aligned}\rho(p) = {} & \frac{\rho_1 n \nu}{\omega_0}\int_0^\infty \sin(\omega_0 t_1)p(t-t_1)dt_1 - \frac{\rho_1 n g \nu^2}{\sqrt{\omega_0^2 - \delta^2}} \\ & \times \int_0^\infty\int_0^\infty\int_0^\infty e^{-\delta t_1}\sin(\sqrt{\omega_0^2-\delta^2}\,t_1)\sin(\omega_0 t_2) \\ & \times \sin(\omega_0 t_3)\cdot p(t-t_1-t_2)\cdot p(t-t_1-t_3)dt_1\,dt_2\,dt_3,\end{aligned} \tag{54}$$

which can be derived from (47)–(49). For example, if the incident pressure varies according to $p_i = p_0\sin(\omega t)$, the nonlinear response, in accordance with the equation of state (54), will be a vibration at the frequency 2ω

$$p^{(2)} = \frac{(g\omega_0^2\nu p_0^2)\cdot(2\delta\omega)\cdot\cos(2\omega t + 2\phi_1 + \phi_2)}{[(\omega_0^2-\omega^2)^2 + (2\delta\omega)^2]\sqrt{(\omega_0^2 - 4\omega^2)^2 + (4\delta\omega)^2}}, \tag{55}$$

where

$$\delta = \frac{1}{2}\nu\rho_1 c_1 n h, \quad \tan\phi_1 = \frac{\omega^2-\omega_0^2}{2\delta\omega}, \quad \tan\phi_2 = \frac{\omega_0^2 - 4\omega^2}{4\delta\omega}. \tag{56}$$

The second example of nonlinear layer illustrating the general approach is the contact between two rough surfaces of solids.[17,18] It can be modeled by a thin layer whose linear and nonlinear elastic moduli are determined by a probabilistic distribution of micro-asperity height of $w(l)$[18]

$$p = \kappa(h_0 - h) + \frac{\beta}{2}(h_0 - h)^2, \tag{57}$$

where

$$\kappa = E\int_{h_0}^{l_0} w(l)\frac{dl}{l}, \quad \beta = E\frac{w(h_0)}{h_0}, \tag{58}$$

where E is Young's modulus of the solid, l_0 is the maximum height of micro-asperity, $h_0(P)$ is the equilibrium thickness

of contact (in the absence of acoustic waves), which depends on the static pressure P applied to the contact. The reflection and transmission factors are dependent on P, and the nonlinear factors have extreme magnitudes at the specific pressure $P=P_*$. Using a measured dependence of the reflected wave on P, one can reconstruct the probability distribution $w(l)$ and evaluate the surface roughness.[18]

For the rough contact (57) and (58), Eq. (37) takes the form

$$\frac{d}{dt}\left(p-\frac{\beta}{2\kappa^2}p^2\right)-\frac{2\kappa}{\rho_1 c_1}p=\frac{2\kappa}{\rho_1 c_1}p_i. \tag{59}$$

An earlier discussed technique for evaluation of surface quality[18] was based on the steady-state response at frequencies ω and 2ω. Equation (59) offers the possibility to calculate the nonlinear pulse response as well.

V. STRONGLY NONLINEAR LAYER

Equation (37) can be solved exactly for several types of nonlinear dependence $f(p)$. Some of these solutions can be related to real inhomogeneities. An important example is

$$\frac{p}{p_*}=\exp(bf)-1, \tag{60}$$

which demonstrates the nonlinear connection between pressure and density (equivalent to a nonlinear stress–strain relationship) typical for strongly loaded solids. At increase in p the function f (proportional to the variation of density of the medium) increases slower than one following a linear law. Such behavior demonstrates a rise in rigidity, showing the nonlinear character of compressibility. Alternatively, an increase in the stretching load leads to fast variation of density: $f\to-\infty$ at p tending to the finite rarefaction $p=-p_*$. The nonlinear dependence (60) can be expanded in power series for weak deviations of pressure and density from the equilibrium state

$$\frac{p}{p_*}=bf+\frac{1}{2}b^2f^2+\frac{1}{6}b^3f^3+\cdots. \tag{61}$$

A comparison between the expansions (61) and (39) demonstrates the meaning of the weak quasilinear deformation constants

$$bp_*=1,\quad b=\frac{2\epsilon}{c_0^2p_0},\quad b^2=\frac{3\chi}{(c_0^2p_0)^2},\cdots. \tag{62}$$

For a periodic signal $p_i(t)=p_0\sin(\omega t)$, the exact steady-state solution of (37) with account for (60) equals

$$\frac{p}{p_*}=\frac{\exp[-a\beta\cos(\omega t)]-S}{\exp[-a\beta\cos(\omega t)]+S}, \tag{63}$$

where

$$S=I_0(a\beta)+2\sum_{n=1}^{\infty}(-1)^nI_n(a\beta)\cos\left(n\omega t-2\arctan\frac{n}{a}\right). \tag{64}$$

Here, I_n are modified Bessel functions

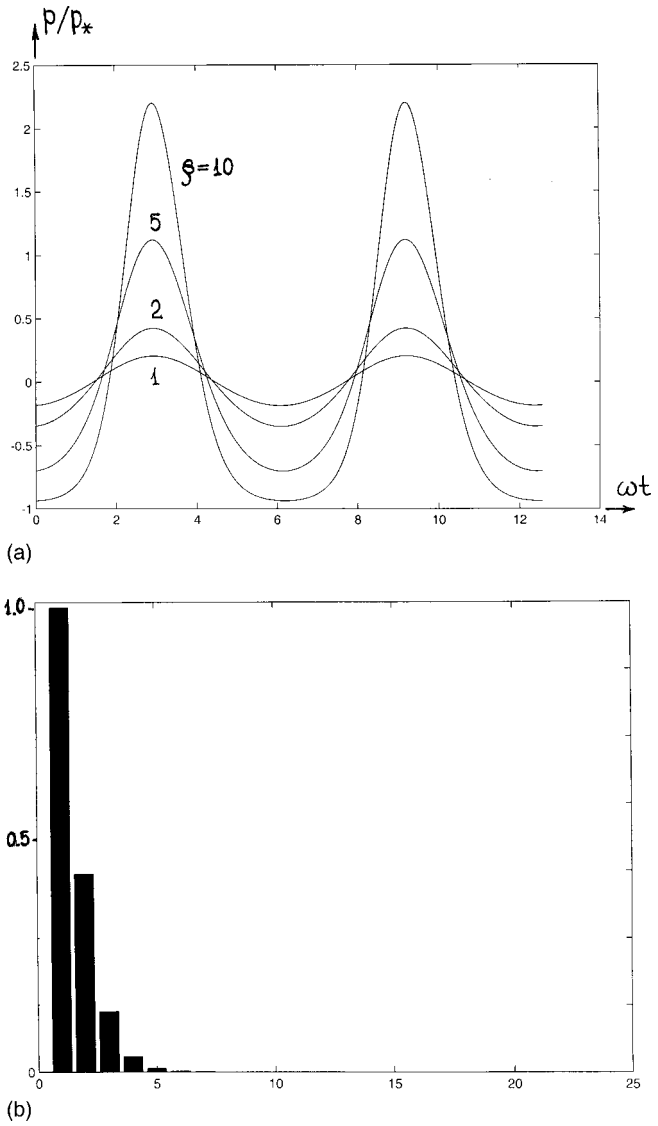


FIG. 5. Strong nonlinear response to harmonic incident signal. Temporal profiles are shown for $a=0.2$, $bp_*=1$ and different values of β (59), (64)—(a). The spectrum corresponding to the curve $\beta=10$ is given in (b).

$$a=\frac{2c_0\zeta}{\omega h}bp_*,\quad \beta=\frac{p_0}{p_*}. \tag{65}$$

The parameter β is a ratio of the amplitude p_0 of the incident wave to the internal pressure p_* of medium characterizing the nonlinear properties of the stress–strain relationship (60). As the parameter β increases, the nonlinear effects must increase as well. Distortion of profile and higher harmonics generation will be more pronounced for rarefaction half-periods, because tensile loads cause fractures at lower magnitudes than compressive loads.

The temporal profiles of strong acoustic response and their corresponding frequency spectra are shown in Fig. 5 and Fig. 6. The numerical values of parameters were $bp_*=1$, $a=0.2$ in Fig. 5, and $a=5$ in Fig. 6. The profiles are presented in Fig. 5(a) and Fig. 6(a) for different β equal to 1, 2, 5, and 10. One can see that increase in both a and β leads to increase in the nonlinear distortion. For $\beta=1$ the profiles in both Fig. 5(a) and Fig. 6(a) are quasiharmonic. With increase in β their amplitude increases, and at $\beta=10$ the shape

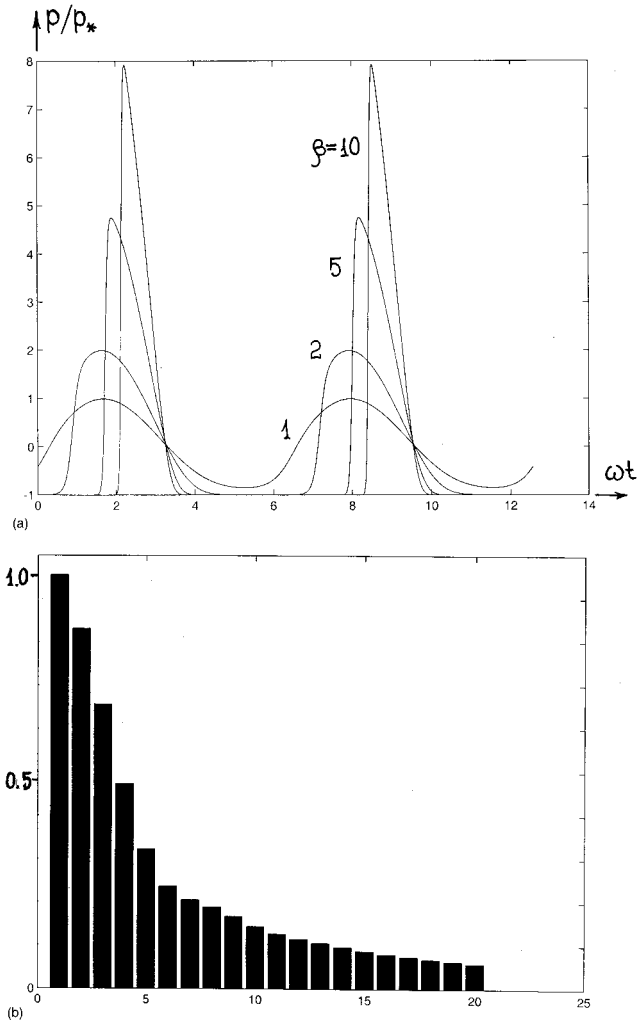


FIG. 6. The same as in Fig. 5, but for $a=5$.

becomes unrecognizable. The sections of positive and negative pressure are distorted quite differently. The section of positive pressure has a structure of a high and sharp peak, and the leading front of it is steep. On the other hand, the section of negative pressure is smoothed. In accordance with the relationship $p(\rho)$ in (60) the normalized negative pressure p/p_* cannot be less than -1.

The spectra in Fig. 5(b) and Fig. 6(b) are calculated for profiles corresponding to the highest nonlinearity, $\beta=10$. In the example with $a=0.2$ [Fig. 5(b)] only 5 harmonics are discernible, but when $a=5$ [Fig. 6(b)] the number of harmonics reaches several tens, and the amplitudes of the second and third ones are comparable in magnitude with the fundamental.

A similar nonlinear behavior of profiles and spectra was observed experimentally[17] for a layer formed by contact of two rough solid boundaries.

Also, the non-steady-state response can be calculated analytically. Let the incident acoustic pressure vary as $p_i = p_0 \cdot \phi(t)$, where ϕ is the function describing the shape of the incident pulse. The solution of the problem (37), (60) can be written as

$$\frac{p}{p_*}=\frac{\exp[a\tau+a\beta\int_{-\infty}^{t}\phi(x)dx]}{C+\int_{-\infty}^{\tau}\exp[ax+a\beta\int_{-\infty}^{x}\phi(x_1)dx_1]dx}-1, \quad (66)$$

where

$$\tau=\frac{t}{t_0}, \quad a=\frac{2c_0t_0\zeta}{h}bp_*. \quad (67)$$

Here, t_0 is a characteristic duration of pulse, and C is an arbitrary constant which must be determined on the basis of the causality principle. Namely, the leading front of response cannot appear earlier than the front of the incident pulse. The solution (66) can be reduced to a simple integral form if the incident pulse has a shape given by

$$\phi=\pm\frac{\tau}{1+\tau^2}\Xi(\tau). \quad (68)$$

Here, Ξ is the Heaviside step function indicating that the disturbance ϕ is identically zero at $\tau<0$. The sign ''plus'' corresponds to a monopolar pulse of positive pressure, and ''minus'' to a negative one. The solution (66), (68) takes the form

$$\frac{p}{p_*}=\frac{(1+\tau^2)^{\pm(1/2)a\beta}e^{a\tau}}{1+a\int_0^{\tau}e^{ax}(1+x^2)^{\pm(1/2)a\beta}dx}-1. \quad (69)$$

The nonlinear pulse response (69) is shown in Fig. 7 and Fig. 8 by solid lines. The corresponding linear pulses are shown by dashed curves for the comparison; their shape was calculated from the formula (23). For all pulses in Fig. 7 the parameter a equals 0.2, and for the ones in Fig. 8 $a=5$.

The other parameter $\beta=p/p_*$ was put to 1, 2.5, and 5 for the three sets of curves at each figure.

One should note the difference in the distortion between positive (a) and negative (b) pulses. An incident rarefaction pulse is deformed more strongly than a compression pulse, and the influence of nonlinear properties on its shape is stronger.

When the incident pulse is very short, in the form of a δ function, Eq. (37) becomes

$$\frac{d}{d\tau}f(p)+ap=ap_0\delta(\tau), \quad a=\frac{2c_0t_0\zeta}{h}, \quad \tau=t/t_0. \quad (70)$$

This can be solved for an arbitrary stress–strain relationship $f(p)$. One can solve the homogeneous equation (70), because its right-hand side equals zero for any moment of time, except at $\tau=0$

$$F(p)\equiv\int\frac{df(p)}{p}=-a\tau+C, \quad p=F^{-1}(-a\tau+C). \quad (71)$$

Following from (70), the constant C must be determined from the condition

$$\int_{-\infty}^{\infty}F^{-1}(-a\tau+C)d\tau=p_0. \quad (72)$$

In particular, for the nonlinear dependence $f=p_*\ln(1+p/p_*)$ [discussed above, see (60)], the general formula (71) reduces to the simple result

$$\frac{p}{p_*}=\frac{(1-e^{-ap_0/p_*})e^{-a\tau}}{1-(1-e^{-ap_0/p_*})e^{-a\tau}}, \quad (73)$$

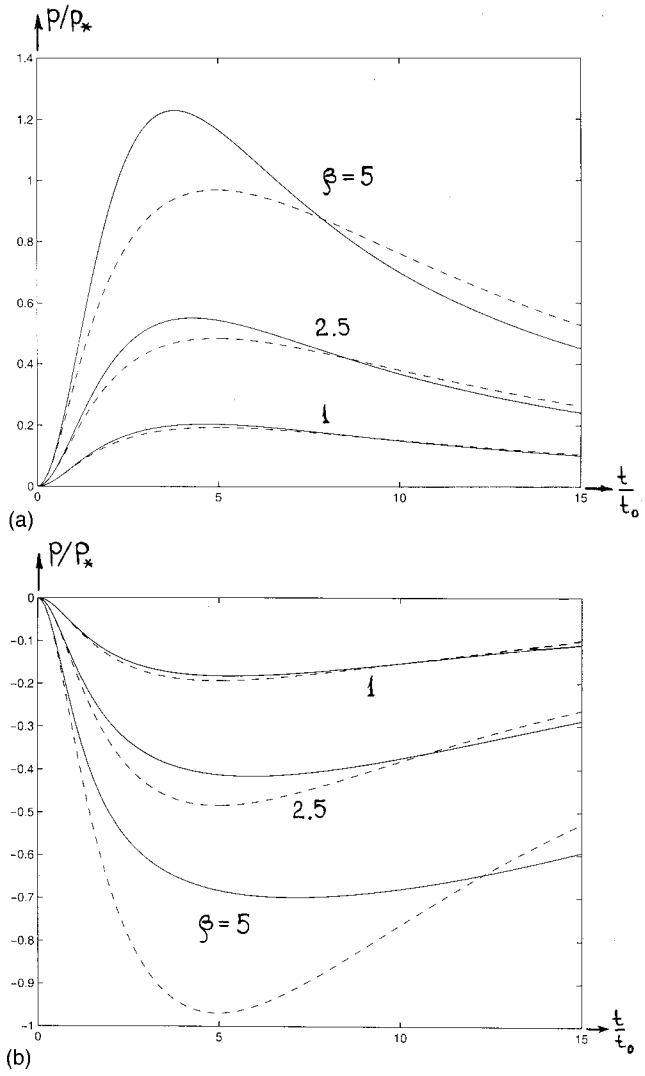


FIG. 7. Strong nonlinear response to a single monopolar incident pulse of compression (a) and rarefaction (b) for $a=0.2$ and different values of β. The corresponding linear response is shown by dashed lines.

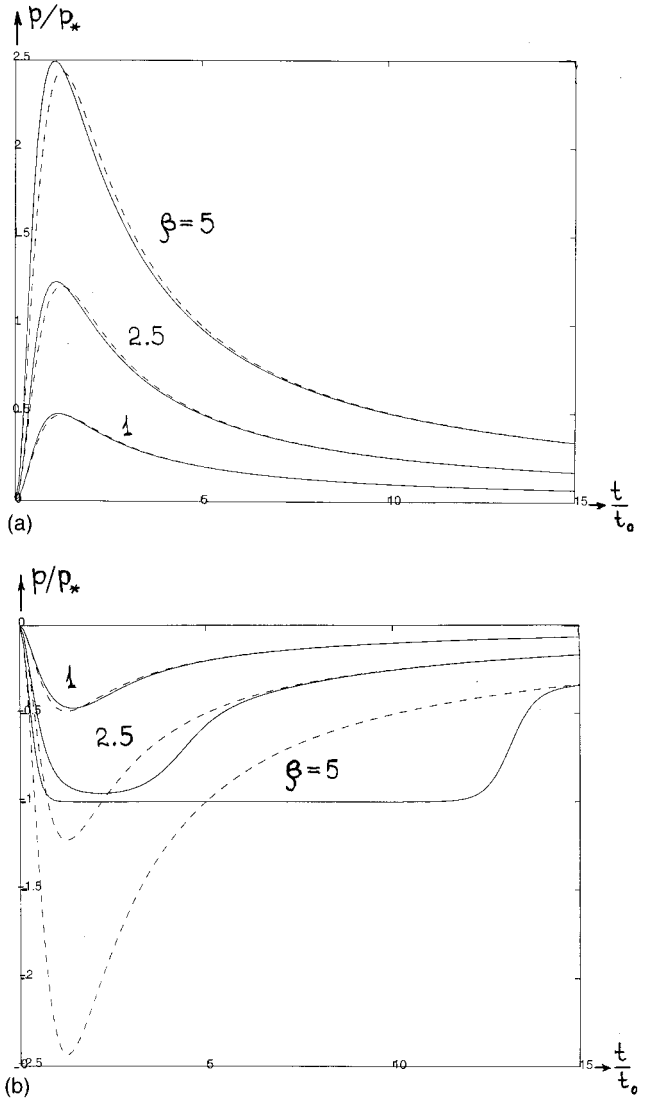


FIG. 8. The same as in Fig. 7, but for $a=5$.

which is valid for $p>0$ and $p_0<0$. The response (73) reaches its extremum

$$\left(\frac{p}{p_*}\right)_{\text{ext}} = -1+\exp\left(a\,\frac{p_0}{p_*}\right), \tag{74}$$

at the instant $\tau=0$ when the incident δ pulse acts on the layer. With increase in time τ the absolute value of the response (73) decreases monotonically.

The temporal profiles of the pulse (73) are shown in Fig. 9 for negative p_0; if $-ap_0/p_*$ is small, $p(\tau)$ is near to the exponent. With increase in $-ap_0/p_*$, the response starts with the value $p/p_*\approx -1$. After a perceptible delay it tends to zero.

Despite that this work is devoted to the direct nonlinear problem, it is pertinent to make some remarks on the possibility of using these results in the solution of inverse nonlinear problems. Many spectral or temporal peculiarities of the nonlinear response can serve as distinguishing characteristics useful for nonlinear nondestructive testing (NNDT). Only the simplest possibilities are discussed here.

The extremal peak pressure (74) can be recorded at the excitation of a layer by a short acoustic pulse. After several (N) measurements, corresponding to different $(p_0)_n$, $n=1,2,3,...,N$, a set of data on peak pressure $(p_{\text{ext}})_n$ can be formed. Consequently, the relationship (74) can be reconstructed with an accuracy depending on the experimental error. But, the dependence (74) is nothing more or less than the dependence $f(p)=ap_0(p)$ used at the derivation of the solution (73). Consequently, this scheme of measurement gives the possibility to evaluate the stress–strain relationship.

A more general scheme follows when considering (70) with its right-hand side corresponding to an arbitrary shape of incident pulse $p_i(\tau)$. It is then convenient to write (70) in the form

$$a(p_i-p)\,\frac{dp}{df}=\frac{dp}{d\tau}. \tag{75}$$

Equation (75) is fulfilled at any moment of τ. Consequently, the derivative dp/df can be considered as a constant, independent of τ. Multiplying (75) by p, it is easy to derive the formula $\langle pp_i\rangle=\langle p^2\rangle$, where the brackets denote the integral of τ over the region where both p and p_i are nonzero. There-

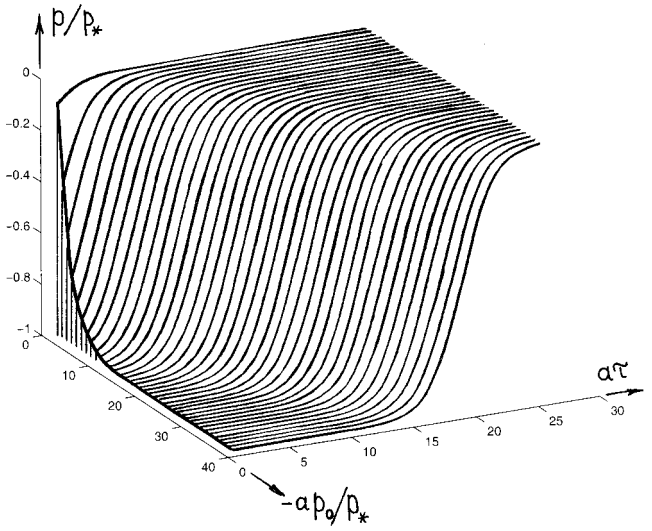


FIG. 9. Strong nonlinear response to a short rarefaction pulse in form of a negative δ function.

after, multiplying (75) by p_i and integrating over $d\tau$ one obtains

$$a\frac{dp}{df} = -\frac{\langle p dp_i/dt\rangle}{\langle p_i^2\rangle - \langle p^2\rangle}. \tag{76}$$

Making a series of measurement at different $(p_0)_n$, the right-hand side Q of (76) can be determined, and (76) takes the form

$$a\left(\frac{dp}{df}\right)_{p=(p_0)_n} = Q[(p_0)_n, a]. \tag{77}$$

During measurements carried out in the linear regime (at small p_0) the right-hand side of (76) does not depend on p_0. This fact makes it possible to determine the constant a. Then, at finite and large values of p_0, after reconstruction of the function Q in a significant number of points, one can integrate (77)

$$a\frac{dp}{Q(p,a)} = df, \tag{78}$$

and restore the stress–strain relationship.

VI. CONCLUSIONS

Much attention has been given in the last few years to nonlinear evaluation of mechanical properties of materials containing mesoscopic inhomogeneities. There have been organized special conferences and special sessions (see, for example, Refs. 19 and 20) devoted to this problem. A great body of significant experimental data has been obtained, and new mechanical models for nonlinear properties of defective solids were developed.

Remote evaluation and testing includes the processing of acoustic signals scattered by linear or nonlinear inhomogeneities. Consequently, the nonlinear response of different internal defects must be studied.

The statement of the problem to which this paper is devoted is most simple: Find the nonlinear pulse response of a nonresonant layer to a plane incident wave. Because of this simplicity, both weak and strong nonlinear responses were studied in detail. These results can be used in qualitative analyses of more complicated scatterers and structures.

Moreover, this investigation line must be continued. First, the nonlinear responses of inhomogeneities having different shapes, internal structures, and physical properties have to be calculated. Second, the results for single inhomogeneities must be used for calculation of responses from ensembles of such scatterers. Third, direct data can be applied in the solution of inverse problems and nonlinear tomography.

The preliminary results of this work were delivered by Hedberg at the conference[20] and published as a brief communication.[21]

ACKNOWLEDGMENTS

Our sincere thanks to P. Johnson, A. Sutin, V. Robsman, I. Solodov, A. E. Sobisevich, and L. E. Sobisevich for fruitful discussions. The work was supported by CRDF, INTAS 99-88, and RFBR grants.

[1] K. A. Naugolnykh and L. A. Ostrovsky, *Nonlinear Wave Processes in Acoustics* (Cambridge University Press, New York, 1998).
[2] O. A. Druzhinin, L. A. Ostrovsky, and A. Prosperetti, "Low-frequency acoustic wave generation in a resonant bubble-layer," J. Acoust. Soc. Am. **100**, 3570–3580 (1996).
[3] L. A. Ostrovsky, A. M. Sutin, I. A. Soutova, A. I. Matveyev, and A. I. Potapov, "Nonlinear, low-frequency sound generation in a bubbly layer: Theory and laboratory experiment," J. Acoust. Soc. Am. **104**, 722–726 (1998).
[4] B. B. Goldberg, J. B. Liu, and F. Forsberg, "Ultrasound contrast agents: A review," Ultrasound Med. Biol. **20**, 4 (1994).
[5] I. Shkolnik, L. Zarembo, and V. Krasilnikov, "On structural nonlinear diagnostics of solids and estimation of their strength," in *Frontiers of Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Elsevier, London, 1990), pp. 589–594.
[6] V. A. Robsman and O. V. Rudenko, "A new method for nonlinear nondestructive testing of developing defects inside inhomogeneous solids," J. Acoust. Soc. Am. **105**, 1015A (1999).
[7] V. A. Robsman, "Nonlinear transformation of noise spectra in the acoustic diagnostics of concrete structures," Sov. Phys. Acoust. **37**, No. 5, 541–543 (1991). "Accumulation and chaotic development of nonlinear acoustical processes in the dynamic loading of geological structures," Acoust. Phys. **39**, 176–183 (1993).
[8] T. Sato, "Industrial and medical application of nonlinear acoustics," in *Frontiers of Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Elsevier, London, 1990), pp. 98–112.
[9] O. V. Rudenko, "Nonlinear methods in acoustic diagnostics (Review)," Russ. J. Nondestruct. Test. **29**, 583–588 (1993).
[10] R. A. Guyer and P. A. Johnson, "Nonlinear mesoscopic elasticity: Evidence for a new class of materials (Review)," in *Physics Today*, April 1999, pp. 30–36.
[11] K. E-A. Van Den Abeele, P. J. Johnson, R. A. Guyer, and K. R. McCall, "On quasi-analytic treatment of hysteretic nonlinear response in elastic wave propagation," J. Acoust. Soc. Am. **101**, 1885–1898 (1997).
[12] Y. Zheng, R. G. Maev, and I. Yu. Solodov, "Nonlinear acoustic applications for material characterization: A review," Can. J. Phys. **77**, 927–967 (1999).
[13] V. A. Burov, I. E. Gurinovich, O. V. Rudenko, and E. Ya. Tagunov, "Reconstruction of the spatial distribution of the nonlinearity parameter and sound velocity in acoustic nonlinear tomography," Acoust. Phys. **40**, 816–823 (1994).
[14] L. M. Brekhovskikh, *Waves in Layered Media*, 2nd ed. (Academic, New York, 1980). L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media* (Springer, New York, 1998).
[15] O. V. Rudenko, and S. I. Soluyan, *Theoretical Foundations of Nonlinear Acoustics* (Plenum, Consultants Bureau, New York, 1977).
[16] V. G. Andreev, V. Yu. Veroman, G. A. Denisov, O. V. Rudenko, and O. A.

Sapozhnikov, ''Nonlinear acoustical aspects of extracorporeal lithotripsy,'' Sov. Phys. Acoust. **38**, 325–328 (1992).
[17] I. Yu. Solodov, ''Ultrasonics of non-linear contacts: Propagation, reflection, and NDE-applications,'' Ultrasonics **36**, 383–390 (1998).
[18] O. V. Rudenko and C. A. Vu, ''Nonlinear acoustic properties of a rough surface contact and acoustodiagnostics of a roughness height distribution,'' Acoust. Phys. **40**, 593–596 (1994).
[19] *Nonlinear Acoustics at the Turn of the Millenium*, Proceedings of the 15th International Symposium on Nonlinear Acoustics, edited by W. Lauterborn and T. Kurz (AIP, Melville, NY, 2000).
[20] Fifth Annual Workshop on Nonlinear Mesoscopic Elasticity, Santa Margherita Ligure, Italy, 4–9 July 2000.
[21] O. V. Rudenko, L. E. Sobisevich, A. L. Sobisevich, and C. M. Hedberg, ''Nonlinear response of a layer to pulse action in diagnostics of small inhomogeneities,'' Dokl. Phys. **45**, 485–488 (2000).

Detection and monitoring of cracks using mechanical impedance of rotor-bearing system

S. Prabhakar, A. S. Sekhar,[a] and A. R. Mohanty
Department of Mechanical Engineering, Indian Institute of Technology, Kharagpur 721302, India

(Received 6 June 2000; accepted for publication 23 August 2001)

The dynamic behavior of structures, in particular, that of a rotor, containing cracks is a subject of considerable current interest. In the present study, the influence of a transverse surface crack, both open and breathing conditions, on the mechanical impedance of the rotor-bearing system, has been investigated by FEM analysis for flexural vibrations. The results show that the mechanical impedance changes substantially due to the presence of crack and follow definite trends with the crack parameters (depth and location) and force location. The normalized mechanical impedance of rotor system is more sensitive to the breathing crack when compared to the open crack. The sudden changes in mechanical impedance have been observed at the multiples of the running frequency for the breathing crack. The measurement of mechanical impedance has been suggested for crack detection and condition monitoring of rotor-bearing systems. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1412447]

PACS numbers: 43.40.Le, 43.40.Kd [PJR]

NOMENCLATURE

[C]	flexibility matrix with crack
[D]	matrix includes gyroscopic effects and damping
D	diameter of shaft
E	modulus of elasticity
e	unbalance eccentricity of the rotor
F	amplitude of impulse force
I_D	diametral moment of inertia
I_P	polar moment of inertia
[K]	stiffness matrix
l	element length
L	length of the rotor
[M]	mass matrix
{q}	vector of nodal quantities
{Q}	force vector
X	location of crack from left bearing
X_f	location of force from left bearing
Z_c	mechanical impedance of the cracked rotor-bearing system
Z_0	mechanical impedance of the rotor-bearing system without crack
Z_c/Z_0	normalized mechanical impedance
α	depth of crack
$\overline{\alpha}$	α/D
ω	operating speed

I. INTRODUCTION

A dynamic analysis of cracked rotors has been a subject of great interest for the last three decades and excellent reviews on this are available.[1–4] One form of damage that can lead to the catastrophic failure if undetected is fatigue cracks. Several techniques have been discussed for crack detection and the monitoring of rotor systems.[1–4] Because of the global nature of the dynamic response of structures, techniques for detecting damages based on the vibration characteristics of structures have been gaining importance. Some researchers[5,6] have suggested the use of natural frequencies and mode shapes while others[7] have recommended the use of frequency response functions for crack detection. The damage due to the presence of cracks can be detected by comparing the frequency changes obtained from the experimental data collected on the structure with the sensitivity of model parameters obtained from an analytical model of the structure. Imam *et al.*[8] presented an on-line rotor crack diagnosis system that will detect even a very small crack by a so-called hystogram signature analysis technique. Some works on the nonlinear vibration of a cracked shaft and structures have also been reported recently.[9,10] Many other crack detection and monitoring techniques were discussed in Ref. 4.

Even though a great deal of achievement in the area of the dynamics of cracked rotors has been obtained, not enough attention has been paid to the studies on the variation of mechanical impedance of the cracked rotors. A nondestructive test method using mechanical impedance was developed by Nezu[11] for detecting damages in the structural member. Recently, Bomnios and Trochids[12] have presented the influence of a transverse crack on the mechanical impedance of the cantilever beam. However, the above techniques[11,12] on the determination of mechanical impedance are applied to structures having open cracks.

In the present paper, the influence of a transverse surface crack, both open and breathing conditions, on the mechanical

[a]Author to whom correspondence should be addressed. Electronic mail: sekhar@mech.iitkgp.ernet.in

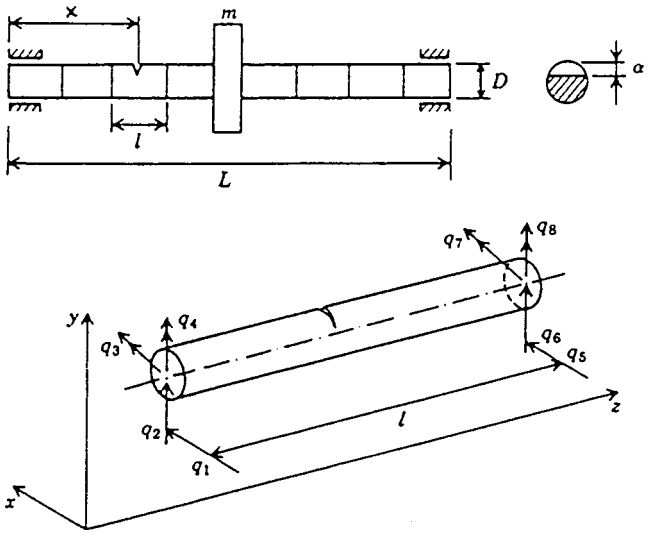


FIG. 1. A model rotor system with a cracked element.

impedance of a rotor-bearing system have been investigated by FEM (finite element method) analysis for flexural vibrations. A parametric study has been done with the crack parameters (crack depth and location) and the location of force in an attempt to use the mechanical impedance change for crack detection and monitoring.

II. SYSTEM EQUATION OF MOTION

Nelson and McVaugh[13] presented a finite element model using Rayleigh beam theory for the rotor-bearing system that consists of rigid disks, distributed parameter finite shaft elements, and discrete bearings. Ozguven and Ozkan[14] presented an extended model by synthesizing the previous models. In the present study we use these models by including the crack model to study the mechanical impedance of the rotor.

The rotor shaft is discretized into finite beam elements, as shown in Fig. 1. Each element has two translational and two rotational degrees of freedom for the bending mode at each node represented by q_1-q_8 (though the element is shown with the transverse crack, the degrees of freedom considered are the same for an uncracked element as well). The

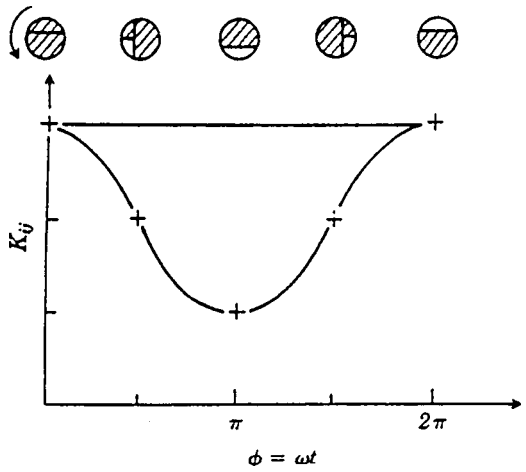


FIG. 2. The breathing crack model.

TABLE I. Rotor-bearing data.

Operating frequency of the rotor, ω	9.54 Hz
Length of the rotor, L	50 cm
Shaft	
Diameter, D	2 cm
Density and modulus of elasticity	7800 kg/m^3, $2.08E\times 11$ N/m^2
Disk	
Location	midspan
Mass, m	5.5 kg
Polar moment of inertia, I_P	0.015 46 kg m^2
Diametral moment of inertia, I_D	0.007 73 kg m^2
Unbalance eccentricity, e	0.1 mm
Crack	
Location, X/L	0–0.5
Depth, $\bar{\alpha}$ (α/D)	0.1–0.4
Bearing (isotropic)	
Stiffness	10^5 N/m
Damping	100 Ns/m
Force (F)	
Magnitude	15 N
Duration	0.01 s
Location, X_f/L	0.04–0.5

discrete bearing stiffness and damping and the rigid disk mass parameters are considered at the corresponding degrees of freedom, after assembling the different shaft elements.

The equation of motion of the complete rotor system in a fixed coordinate system can be written as

$$[\mathrm{M}]\{\ddot{q}\}+[\mathrm{D}]\{\dot{q}\}+[\mathrm{K}]\{\mathrm{q}\}=\{\mathrm{Q}\}, \tag{1}$$

where the mass matrix includes the rotary and translational mass matrices of the shaft and the rigid disk mass, and the diametral moment of inertia. The matrix [D] includes the gyroscopic moments and the bearing damping. The stiffness matrix considers the stiffness of the shaft, the bearing, and the cracked shaft element. The details of the individual matrices of Eq. (1), except for the cracked element, are given in Refs. 13–15.

The excitation matrix {Q} in Eq. (1) includes the unbalance force components in the x and y directions and the external impulse force of constant magnitude in the y direction at the corresponding degrees of freedom. A disk of mass M, having an unbalance eccentricity e, is considered. The unbalance force components in the x and y directions are given as

$$F_x = \mathrm{Me}\,\omega^2 \cos \omega t, \tag{2}$$

$$F_y = \mathrm{Me}\,\omega^2 \sin \omega t. \tag{3}$$

TABLE II. Eigenfrequencies of an undamped and uncracked rotor-bearing system.

Mode	Eigenfrequency (Hz)
1	24.27
2	95.20
3	175.65

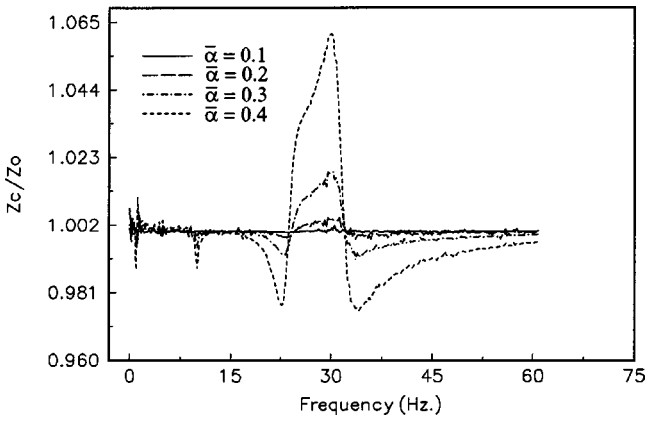


FIG. 3. Variation of impedance with frequency for different crack depths; $X_f/L=0.04$; $X/L=0.43$.

III. MECHANICAL IMPEDANCE

Mechanical impedance $Z(X_f)$ of the rotor-bearing system at any force location X_f from the left bearing location is defined as the ratio of force $F(X_f)$ to the velocity response $V(X_f)$. Mathematically,

$$Z(X_f)=F(X_f)/V(X_f). \tag{4}$$

In the present paper, an external impulse force of constant magnitude for a short duration of time (0.01 s) has been applied on the rotor-bearing system with unbalance. The Houbolt time marching technique[16] has been used to obtain the steady-state velocity response in the time domain. Although the stiffness varies with the time, as explained later, when the shaft is cracked during a small time step, the stiffness was assumed to be constant and the same time marching scheme was used.

IV. CRACK MODELING

In the present study the transverse crack has been considered for both open and breathing conditions. The flexibility matrices of the cracked section for flexural vibrations as given in Papadopoulos and Dimarogonas[17] and utilized in the FEM analysis of Sekhar and Prabhu[15,18] have been used for the crack modeling, and the details are presented here.

A. Open crack

According to the Saint–Venant principle, the stress field is affected in the region adjacent to the crack. The element stiffness matrix, except for the cracked element, may be regarded as unchanged under certain limitations of the element

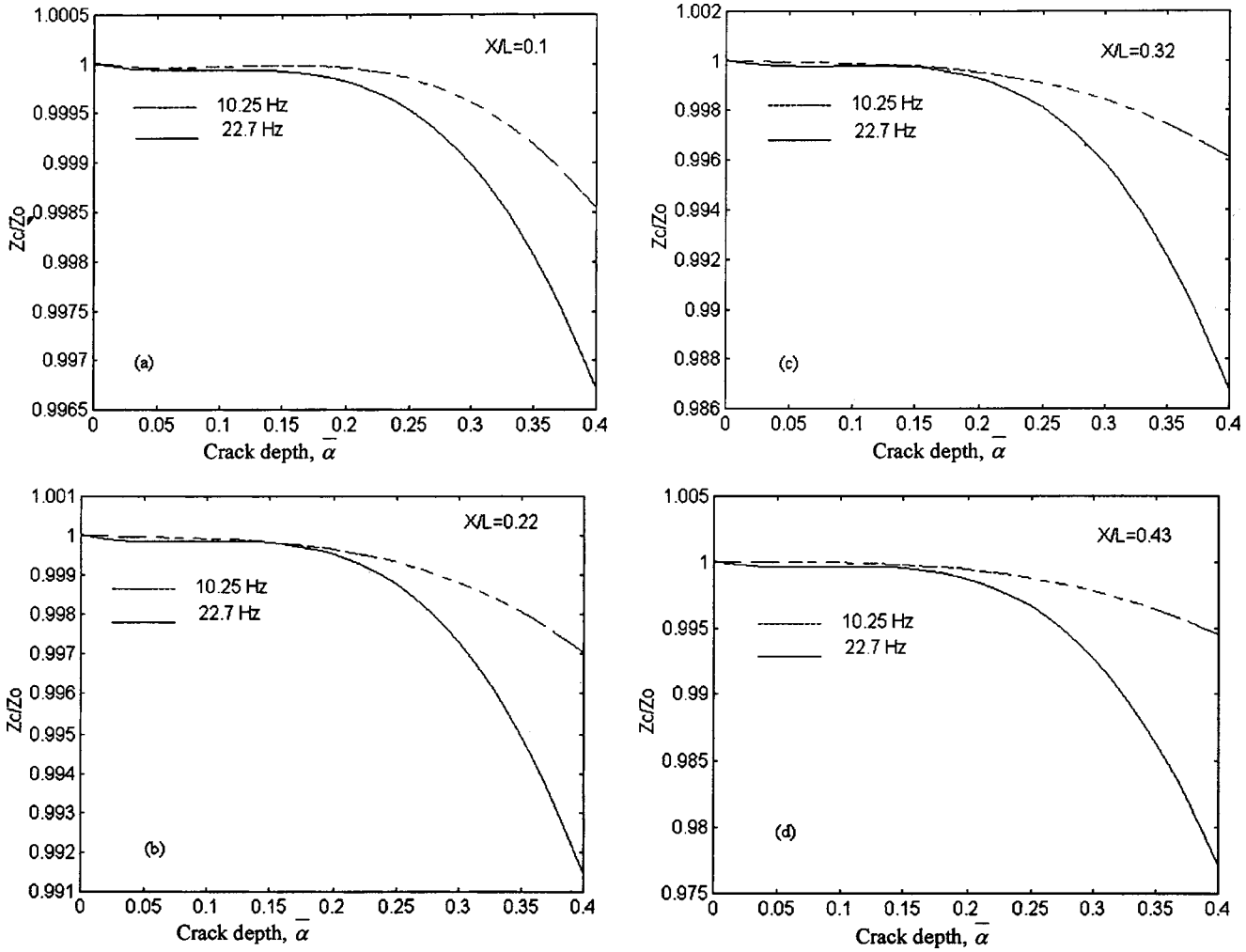


FIG. 4. Variation of impedance with crack parameters at different frequencies; $X_f/L=0.04$. (a) For $X/L=0.1$; (b) for $X/L=0.22$; (c) for $X/L=0.32$; and (d) for $X/L=0.43$.

size.[19] The additional strain energy of a crack has been studied in fracture mechanics, and the flexibility coefficient expressed by the stress intensity factor can easily be derived by means of Castigliano's theorem in the linear elastic range. The shaft is discretized into several finite elements. The behavior of the elements situated to the right of the cracked element may be regarded as external forces applied to the cracked element, while the behavior of elements situated to its left may be regarded as constraints.[19] Thus, the flexibility matrix of the cracked element with constraints is calculated.

Upon neglecting the shearing action and using the strain energy, the flexibility coefficients for an element as shown in Fig. 1, without a crack, are found to be given by

$$C_0=\begin{bmatrix} l^3/3EI & & \text{SYM} & \\ 0 & l^3/3EI & & \\ 0 & -l^2/2EI & l/EI & \\ l^2/2EI & 0 & 0 & l/EI \end{bmatrix},$$

where EI is the bending stiffness and l is the element length. The additional strain energy due to the crack results in a local flexibility matrix,

$$C_c=\frac{1}{\overline{F_0}}\begin{bmatrix} \overline{C}_{11}R & & \text{SYM} & \\ 0 & \overline{C}_{22}R & & \\ 0 & 0 & \overline{C}_{33}/R & \\ 0 & 0 & \overline{C}_{43}/R & \overline{C}_{44}/R \end{bmatrix},$$

where $F_0=\pi ER^2/(1-\nu^2)$, $R=D/2$, and $\nu=0.3$.

The dimensionless compliance coefficients, $\overline{c}_{ij}$, are computed from the derivations discussed in Ref. 17. The total flexibility matrix for the cracked section is given as $[C]=[C_0]+[C_c]$.

From the equilibrium condition

$$(q_1,q_2,\ldots,q_8)^T=[T](q_5,\ldots,q_8)^T, \tag{5}$$

where the transformation is

$$T=\begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & l & -1 & 0 \\ -l & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.,$$

The stiffness matrix of the cracked element is written as

$$[K_c]=[T][C]^{-1}[T]^T. \tag{6}$$

B. Breathing crack

The breathing action of the crack, i.e., its opening and closing as represented by Papadopoulos and Dimarogonas[20] is illustrated in Fig. 2. During the shaft's rotation, the crack opens and closes, depending on the rotor deflection. For the large class of machines, the static deflection is much greater than the rotor vibration. With this assumption the crack is closed when $\phi=0$ and it is fully open when $\phi=\pi$.

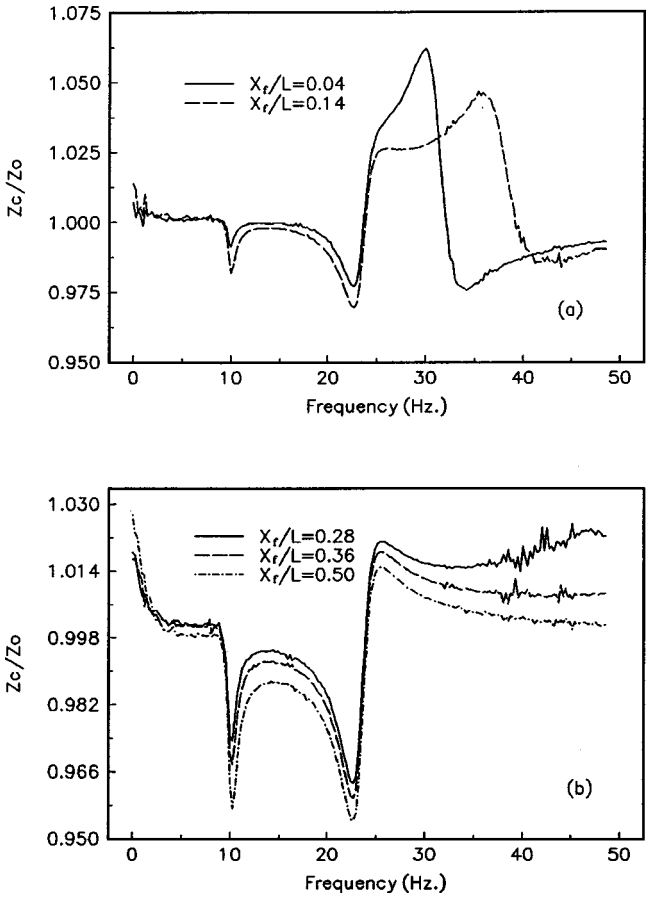


FIG. 5. Variation of impedance with frequency for different force locations; $\overline{\alpha}=0.4$; $X/L=0.43$. (a) For $X_f/L=0.04$ and 0.14; (b) for $X_f/L=0.28$, 0.36, and 0.5.

The transverse surface crack on the shaft element introduces considerable local flexibility due to strain energy concentration in the vicinity of the crack tip under load. The additional strain energy due to the crack results in a local flexibility matrix C_c, which will be C_{op} and C_{HC} for a fully open crack and half-open, half-closed conditions, respectively:

$$C_{\text{op}}=\frac{1}{\overline{F_0}}\begin{bmatrix} \overline{C}_{11}R & & \text{SYM} & \\ 0 & \overline{C}_{22}R & & \\ 0 & 0 & \overline{C}_{33}/R & \\ 0 & 0 & \overline{C}_{43}/R & \overline{C}_{44}/R \end{bmatrix},$$

$$C_{\text{HC}}=\frac{1}{2F_0}\begin{bmatrix} \overline{C}_{22}R & & \text{SYM} & \\ 0 & \overline{C}_{11}R & & \\ 0 & 0 & \overline{C}_{44}/R & \\ 0 & 0 & \overline{C}_{34}/R & \overline{C}_{33}/R \end{bmatrix},$$

where $F_0=\pi ER^2/(1-\nu^2)$, $R=D/2$, and $\nu=0.3$.

Again the dimensionless compliance coefficients, $\overline{c}_{ij}$, are computed from the derivations discussed in Ref. 17. The total flexibility matrix for the cracked section is given as $[C]=[C_0]+[C_c]$.

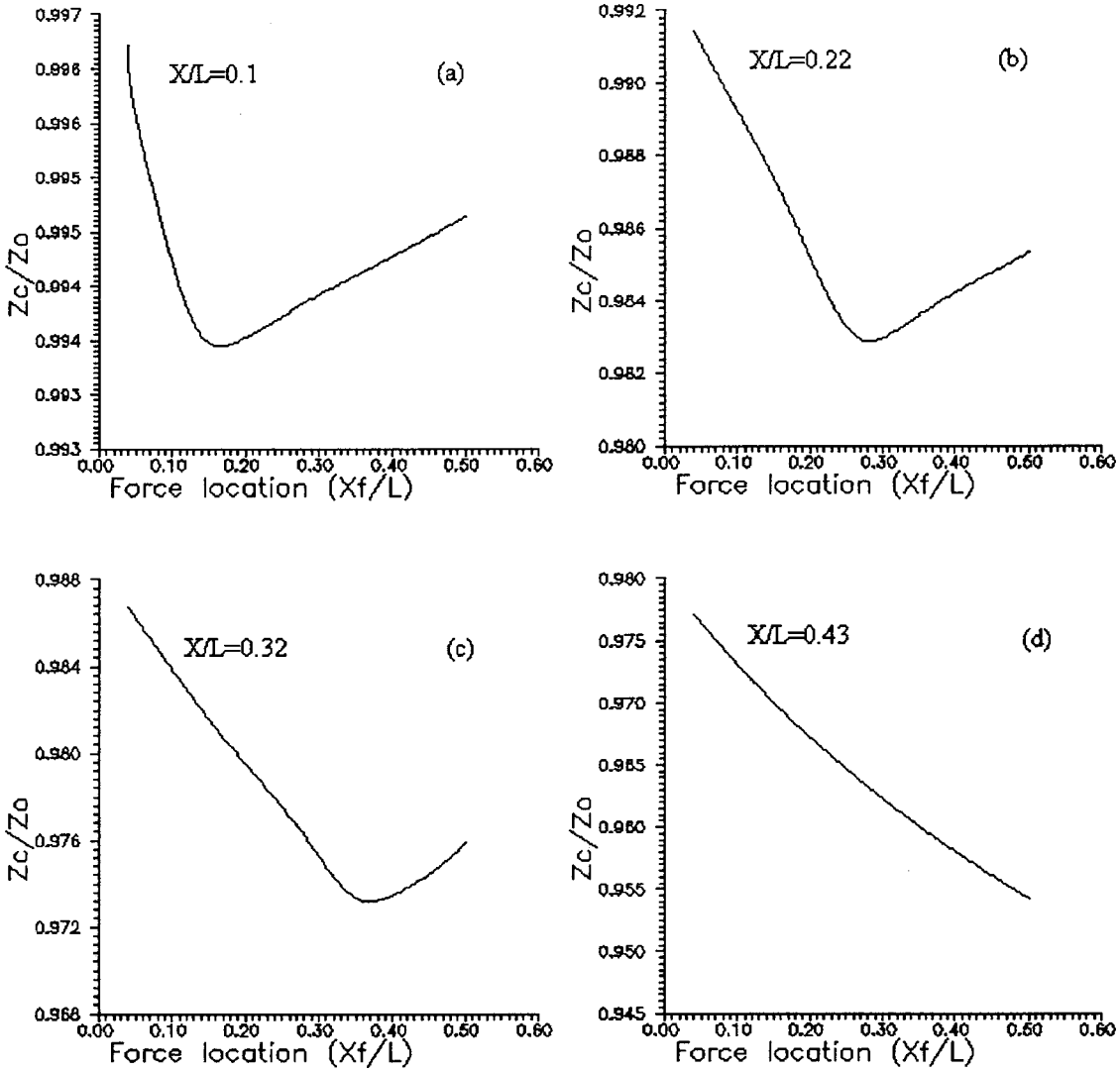


FIG. 6. Variation of impedance with force location; frequency=22.7 Hz.; $\bar{\alpha}=0.4$. (a) For $X/L=0.1$; (b) for $X/L=0.22$; (c) for $X/L=0.32$; (d) for $X/L=0.43$.

From the equilibrium condition

$$(q_1,q_2,\ldots,q_8)^T=[\mathrm{T}](q_5,\ldots,q_8)^T. \tag{7}$$

The matrices $[\mathrm{C}_0]$ and $[\mathrm{T}]$ are same for open and breathing crack conditions. The stiffness of the cracked element is given in Eq. (6). When the crack is assumed to affect only the stiffness, the stiffness matrix of the cracked element, $[\mathrm{K}_c]$, replaces the stiffness matrix of the element that was originally uncracked, when assembling the stiffness matrix of the shaft, for $[\mathrm{K}]$ in Eq. (1).

When the shaft is cracked, during the rotation the stiffness varies with time, or with angle. The variation may be expressed by a truncated cosine series,[20]

$$[\mathrm{K}]=[\mathrm{K}(\omega t)]=[\mathrm{K}_0]+[\mathrm{K}_1]\cos\omega t+[\mathrm{K}_2]\cos 2\omega t + [\mathrm{K}_3]\cos 3\omega t+[\mathrm{K}_4]\cos 4\omega t, \tag{8}$$

where $[\mathrm{K}_\eta]$, $\eta=0,1,\ldots,4$, are fitting coefficient matrices, determined from the known behavior of the stiffness matrix at certain angular locations, as explained by Papadopoulos and Dimarogonas.[20] Specifically,

$$[\mathrm{K}]=[\mathrm{K}]_{\phi=0} \quad \text{and} \quad (\partial^2/\partial\phi^2)[\mathrm{K}]=0, \quad \text{at} \quad \phi=\omega t=0,$$

$$[\mathrm{K}]=[\mathrm{K}]_{\phi=\pi/2}, \quad \text{at} \quad \phi=\pi/2, \tag{9}$$

$$[\mathrm{K}]=[\mathrm{K}]_{\phi=\pi} \quad \text{and} \quad (\partial^2/\partial\phi^2)[\mathrm{K}]=0, \quad \text{at} \quad \phi=\omega t=\pi.$$

$[\mathrm{K}]_{\phi=0}$, $[\mathrm{K}]_{\phi=\pi/2}$, and $[\mathrm{K}]_{\phi=\pi}$ are the stiffness matrices, $\mathrm{K_{UC}}$, $\mathrm{K_{op}}$, and $\mathrm{K_{HC}}$ for the uncracked, fully open crack, and half-open, half-closed cracked conditions, respectively. These are obtained from the compliance matrices C_0, C_{op}, and C_{HC}, respectively.

The application of conditions (9) to Eq. (8) yields the matrices $[\mathrm{K}_\eta]$, $\eta=0,1,2,3,4$, element by element as solutions of the 5×5 linear system of algebraic equations. The constant matrices of Eq. (8) are

$$\begin{aligned}
[\mathrm{K}_0]&=(5K_{\mathrm{op}}+5K_{\mathrm{UC}}+6K_{\mathrm{HC}})/16,\\
[K_1]&=9(K_{\mathrm{UC}}-K_{\mathrm{op}})/16,\\
[\mathrm{K}_2]&=(K_{\mathrm{op}}+K_{\mathrm{UC}}-2K_{\mathrm{HC}})/4,\\
[\mathrm{K}_3]&=(K_{\mathrm{op}}-K_{\mathrm{UC}})/16,\\
[\mathrm{K}_4]&=(-K_{\mathrm{op}}-K_{\mathrm{UC}}+2K_{\mathrm{HC}})/16.
\end{aligned} \tag{10}$$

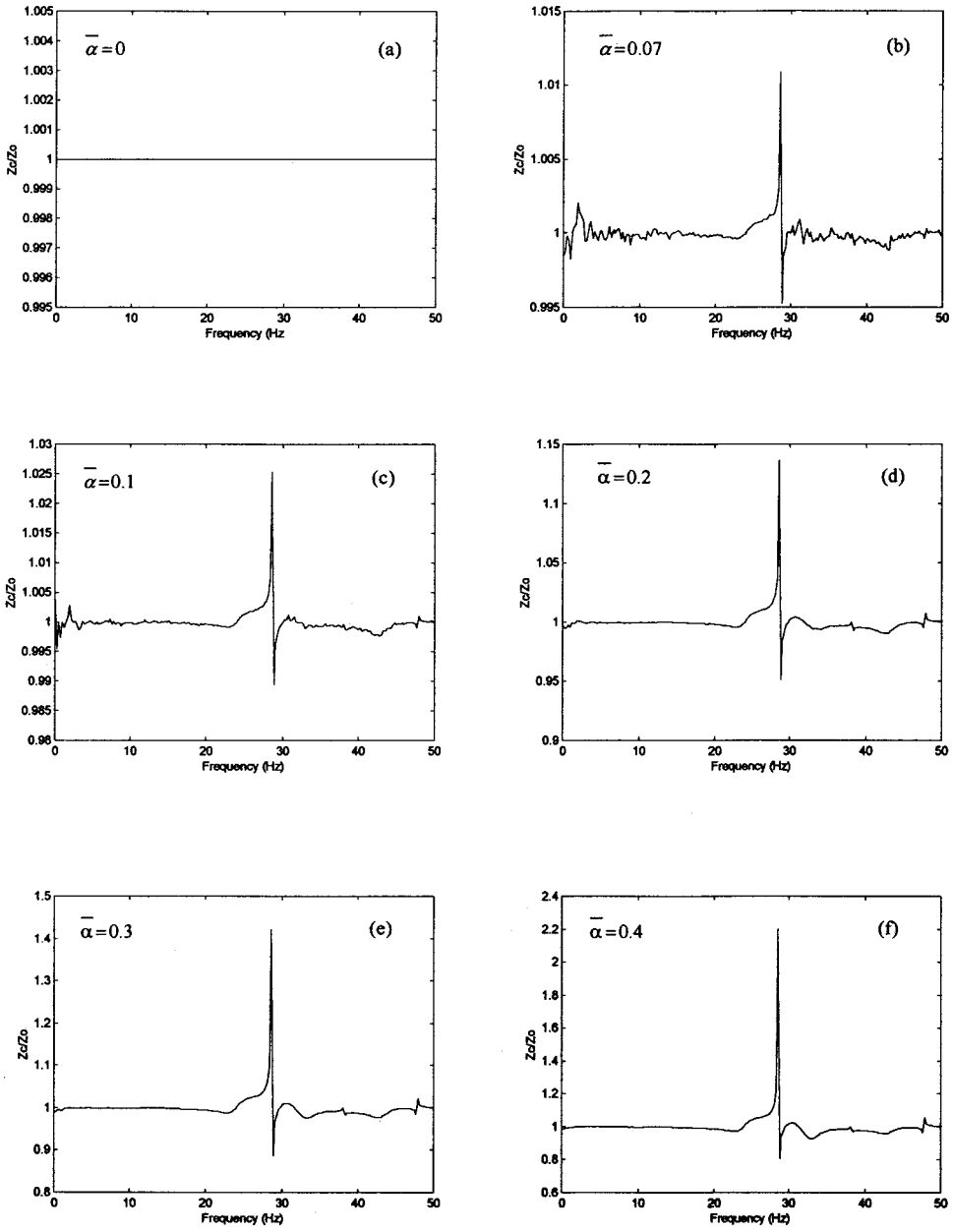


FIG. 7. Variation of impedance with frequency for different crack depths; $X_f/L=0.04$; $X/L=0.43$. (a) For $\bar{\alpha}=0$; (b) for $\bar{\alpha}=0.07$; (c) for $\bar{\alpha}=0.1$; (d) for $\bar{\alpha}=0.2$; (e) for $\bar{\alpha}=0.3$; (f) for $\bar{\alpha}=0.4$.

V. RESULTS AND DISCUSSION

The rotor shaft is discretized into ten finite elements. The stiffness of the cracked element replaces the uncracked element while assembling the stiffness matrix of the rotor-bearing system. The data used for the rotor-bearing system with the crack are shown in Table I. The eigenfrequencies for undamped and uncracked rotor-bearing system are shown in Table II. An impulse force of constant magnitude has been applied for short duration of time ($t=0.01$ s) at different locations of the rotor-bearing system. The mechanical impedance has been found and plotted with frequency [obtained from the fast Fourier transform (FFT)] for the rotor-bearing system with and without cracks. A parametric study has been done for the crack depth, the crack location, and the force location for open and breathing cracks.

A. Open crack

The normalized mechanical impedance (Z_c/Z_0) has been considered for the parametric study. Figure 3 shows the variation of mechanical impedance with frequency for different crack depths by applying an impulse force near the bearing point. It is clear from Fig. 3 that the mechanical impedance changes substantially with the presence of crack. It is well known that the natural frequencies decrease due to the presence of a crack. This can be observed from Fig. 3, as the crack depth increases, the impedance peaks to a maximum near the first eigenfrequency (24.7 Hz) and shifts slightly toward lower values of frequencies. The effects of crack depth and location on the mechanical impedance at the frequencies near the operating speed (10.25 Hz) and near the critical speed (22.7 Hz) are shown in Figs. 4(a)–(d). The impedance decreases as the crack depth increases. Also for the same crack depth, the impedance decreases more as the location of the crack is toward the disk or the crack is situated away from the point of application of force, as compared to that of the other locations of the crack.

Even though it is practically difficult to apply force at a location other than the bearing point, numerical simulations

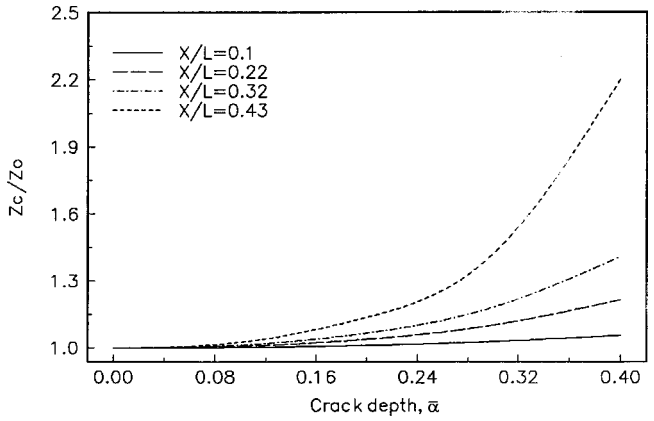


FIG. 8. Variation of impedance at 28.56 Hz with crack depth and location; $X_f/L=0.04$.

are presented here to investigate the effect of force location on mechanical impedance and are shown in Figs. 5(a) and (b) and 6(a)–(d). It is observed that the impedance changes suddenly at the operating frequency and at the first eigenfrequency [see Figs. 5(a) and (b)] and between these frequencies the impedance decreases with crack depth. Another important observation from Fig. 6 is that the impedance decreases until the force application point crosses the crack and then increases gradually as the force moves toward the disk. From these results, it can be noted that the change in the mechanical impedance can be used for possible crack detection and the monitoring of rotor systems.

B. Breathing crack

The parametric study has been repeated for the same rotor system by considering the crack to be breathing. The variations of impedances with a frequency for different crack depths by keeping crack and force locations unchanged are shown in Figs. 7(a)–(f). Even though the change of impedance has been observed over the range of frequencies, the sudden changes of mechanical impedance are evident from the figures at the multiples of operating frequency (9.54 Hz) and prominently at three times the operating frequency (3ω). This is happening due to the breathing action of the crack. It is noted here that the mechanical impedance is sensitive, even for low crack depths, $\bar{\alpha}<0.1$ [see Fig. 7(b)], where there is a significant change in the impedance as compared to the no crack ($\bar{\alpha}=0$) condition. The results suggest that the measurements of mechanical impedance can be used for crack detection and monitoring the crack. The present method uses a simple model. Also, since it is linear, on-line crack condition monitoring is relatively easier as compared to nonlinear models.[9,10]

Figure 8 shows the variation of mechanical impedance at the frequency of 28.56 Hz (3ω) with crack depth for different crack locations. Mechanical impedance increases as the crack depth increases at 28.56 Hz frequency (see Fig. 8), which is unlike the open crack case (Fig. 4), since the plot is drawn at a different frequency.

The effects of force location on impedance are shown in Figs. 9(a)–(d) and Figs. 10(a)–(d) similar to the open crack case. At all the locations, the breathing action of the crack can be clearly identified by the sudden change of mechanical impedance at the multiples of the running frequency (see Fig. 9). The mechanical impedance changes substantially over the range of frequencies as the force location moves toward the disk and the crack is near the disk. The trend is same as for an open crack i.e., the impedance decreases until

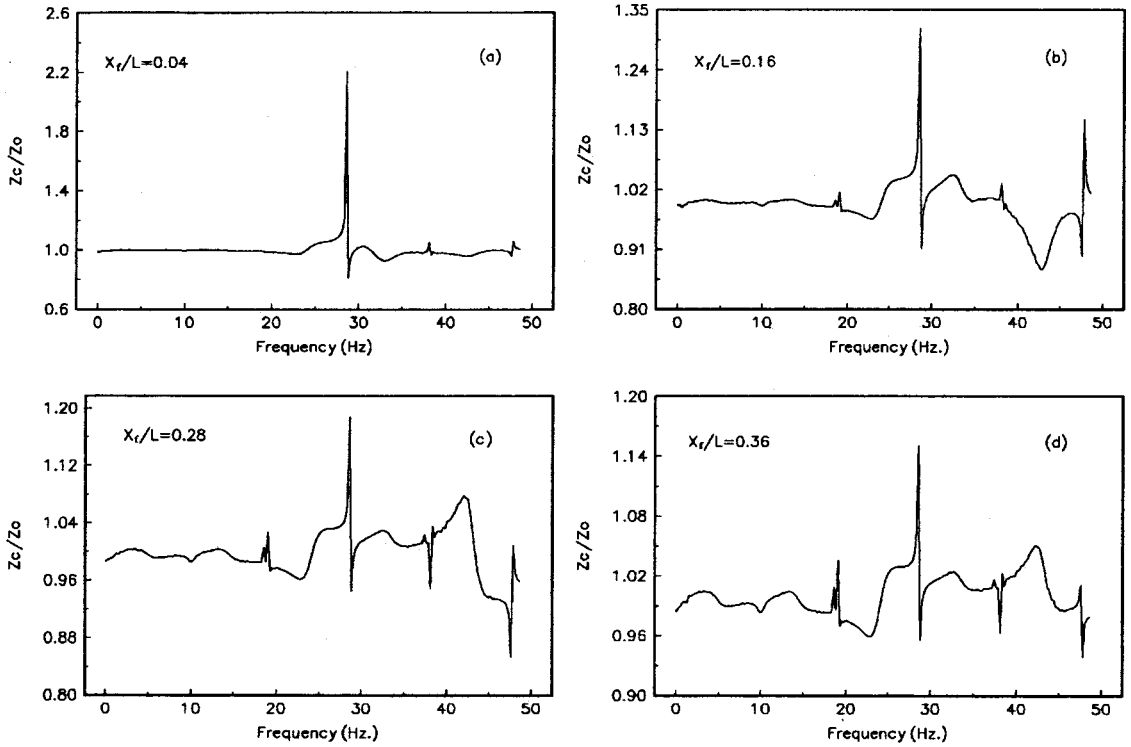


FIG. 9. Variation of impedance with frequency for different force locations; $\bar{\alpha}=0.4$; $X/L=0.43$. (a) For $X_f/L=0.04$; (b) For $X_f/L=0.14$; (c) for $X_f/L=0.28$; (d) for $X_f/L=0.36$.

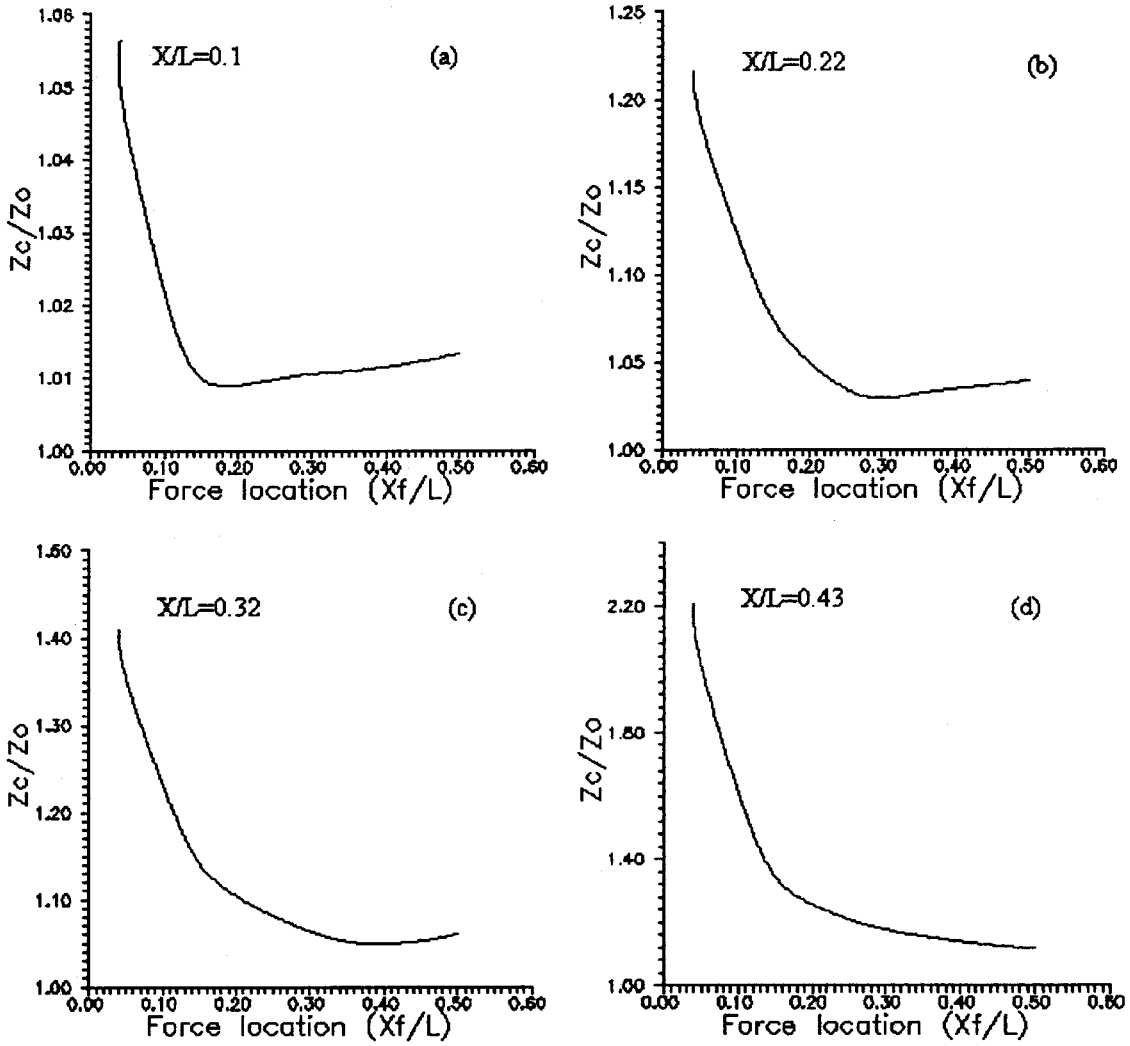


FIG. 10. Variation of impedance with force location; frequency=28.56 Hz; $\bar{\alpha}=0.4$. (a) For $X/L=0.1$; (b) for $X/L=0.22$; (c) for $X/L=0.32$; (d) for $X/L=0.43$.

the force application point crosses the crack and then increases gradually as the force moves toward the disk. However, the normalized mechanical impedance change is highly sensitive to depth in the case of a breathing crack, which is clear from Fig. 7 as compared to Fig. 3. As observed for structures, the mechanical impedance of the rotor system could be used as an additional crack information carrier since it depends on the position at which it is measured.[12]

C. Crack prediction scheme

An exhaustive parametric study can be made similar to the results obtained in the present paper to develop a nomogram for the detection of cracks in a rotor-bearing system. Using such a nomogram with crack depth, crack location, impedance change, and significant frequencies as parameters, following a step-by-step prediction scheme is proposed for detecting cracks in a rotor-bearing system.

(i) Identify the significant frequencies, which are sensitive to an impedance change with crack depth and location. These frequencies are the one near the natural frequency of the rotor system and the other near the operating frequency in case of open cracks, and multiples of operating frequency in case of breathing cracks.

(ii) Quantify the impedance change at these significant frequencies.

(iii) Based on (ii), and with the help of the nomogram it is possible to find the crack depth and location in a rotor system.

VI. CONCLUSIONS

The rotor-bearing system has been modeled using FEM. The influence of a transverse crack on the mechanical impedance of a rotor-bearing system has been analyzed for both open and breathing cracks.

The normalized mechanical impedance changes substantially with the presence of a crack exhibiting definite trends depending upon the location and depth of the crack. The impedance decreases as the crack depth increases at the significant frequencies for an open crack. For the same crack depth, the impedance decreases more as the crack moves toward the disk or the crack moves away from the point of application of the force.

The impedance changes suddenly at the multiples of running frequency for a breathing crack. The effect of the force location on normalized impedance (Z_c/Z_0) is the same for both open and breathing cracks, but a breathing crack is highly sensitive to mechanical impedance compared to an open crack.

It has been shown that the changes in rotor impedance due to a crack are substantial, as compared to that of natural frequencies of the rotor bearing system. The results suggest that the measurements of mechanical impedance can be used for crack detection and monitoring the crack.

[1] A. D. Dimarogonas and S. A. Paipetis, *Analytical Methods in Rotor Dynamics* (Applied Science, London, 1983), pp. 144–193.
[2] J. Wauer, ''Dynamics of cracked rotors, Literature survey,'' Appl. Mech. Rev. **43**, 13–17 (1990).
[3] R. Gaush, ''A survey of the dynamic behavior of simple rotating shaft with a transverse crack,'' J. Sound Vib. **160**, 313–332 (1993).
[4] A. D. Dimarogonas, ''Vibration of cracked structures: A state of the art review,'' Eng. Fract. Mech. **55**, 831–857 (1996).
[5] P. F. Rizos, N. Aspragathos, and A. D. Dimarogonas, ''Identification of crack location and magnitude in a cantilever beam from vibration modes,'' J. Sound Vib. **138**, 381–388 (1990).
[6] A. K. Panday, M. Biswas, and M. M. Samman, ''Damage detection from changes in curvature mode shapes,'' J. Sound Vib. **145**, 321–332 (1991).
[7] M. A. Mannan and M. H. Richardson, ''Detection and location of structural cracks using FRF measurements,'' *Proceedings of the 8th International Model Analysis Conference*, 1990, Vol. 1, pp. 652–657.
[8] I. Imam, S. H. Azzaro, R. J. Bankert, and J. Scheibel, ''Development of on-line rotor crack detection and monitoring system,'' Trans. ASME, J. Vib., Acoust., Stress, Reliab. Des. **111**, 241–250 (1989).
[9] I. Ballo, ''Nonlinear effects of vibration of a continuous transverse cracked slender shaft,'' J. Sound Vib. **217**, 321–333 (1998).
[10] J. Brandon, ''Nonlinear vibration of cracked structures: perspectives and horizons,'' Shock Vib. Dig. **32**, 273–280 (2000).
[11] K. Nezu and H. Kidoguchi, ''A new damage detecting method by mechanical impedance measurements,'' Bull. JSME **23**, 2125–2131 (1980).
[12] G. Bamnios and A. Trochidis, ''Mechanical impedance of a cracked cantilever beam,'' J. Acoust. Soc. Am. **97**, 3625–3635 (1995).
[13] H. D. Nelson and J. M. McVaugh, ''The dynamics of rotor-bearing systems using finite elements,'' ASME J. Eng. Ind. **98**, 593–600 (1976).
[14] H. N. Ozguven and Z. L. Ozkan, ''Whirl speeds and unbalance response of multi bearing rotor using finite elements,'' Trans. ASME, J. Vib., Acoust., Stress, Reliab. Des. **106**, 72–79 (1984).
[15] I. Masayuki and Y. Hiroshi, ''A general method of analysis for dynamic problem of rotor systems,'' Bull. JSME **27**, 1716–1722 (1984).
[16] A. S. Sekhar and B. S. Prabhu, ''Transient analysis of a cracked rotor passing through the critical speed,'' J. Sound Vib. **173**, 415–421 (1994).
[17] C. A. Papadopoulos and A. D. Dimarogonas, ''Coupled longitudinal and bending vibrations with an open crack,'' J. Sound Vib. **117**, 81–93 (1987).
[18] A. S. Sekhar and B. S. Prabhu, ''Crack detection and vibration characteristics of cracked shafts,'' J. Sound Vib. **157**, 375–381 (1992).
[19] G. L. Qian, S. N. Gu, and J. S. Jiang, ''The dynamic behavior and crack detection of a beam with a crack,'' J. Sound Vib. **138**, 233–243 (1990).
[20] C. A. Papadopoulos and A. D. Dimarogonas, ''Stability of cracked rotors in the coupled vibration mode,'' Trans. ASME, J. Vib., Acoust., Stress, Reliab. Des. **110**, 356–359 (1988).

An experimental study of the impedance of perforated plates with grazing flow

N. S. Dickey and A. Selamet
Department of Mechanical Engineering, The Ohio State University, 206 West 18th Avenue, Columbus, Ohio 43210-1107

M. S. Ciray
Advanced Engineering, ArvinMeritor Exhaust Division, Columbus, Indiana 47201

(Received 16 August 2000; revised 2 May 2001; accepted 12 July 2001)

An experimental investigation of the linear impedance of perforated interfaces exposed to grazing fluid flow is presented, including a description of the branch-type setup used in the study. The experimental setup employs a movable microphone in the branch duct that improves the error characteristics of the system. The impedance parameters of three perforated samples have been measured including one circular, square-edged orifice and two production muffler louver geometries. The data are presented in terms of the equivalent length and resistance of the samples, with the friction velocity being used to characterize the grazing flow. At higher grazing flow velocities, the equivalent length of each sample decreased by an amount representative of the Rayleigh end correction for one side of an orifice. Resistance for the louvers appears to be somewhat less sensitive to grazing flow than the circular orifice, though all of the samples showed significant increases. While the experimental results for all three samples showed similar basic trends, distinct and substantial differences were observed between each perforate geometry. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1400736]

PACS numbers: 43.50.Gf, 43.58.Bh [MRS]

I. INTRODUCTION

Perforated interfaces are commonly used in attenuation devices for duct systems. Typical applications include silencers for ventilation systems, wall treatments for aircraft engines, and silencers for the breathing systems of reciprocating engines and compressors. Of particular interest in this study is the use of perforated tubes in automotive silencers. The exhaust system muffler in a passenger car, for example, often includes several perforated tube sections.[1,2] The perforated sections generally consist of a number of small perforations that are spaced rather closely together. Geometries that are used for the individual perforations can vary, but circular orifices and stamped louvers are common.

In an automotive silencer, the flows in and near the perforations will have steady components in addition to the acoustic oscillations. These mean flows have a significant effect on the acoustic behavior of the perforations. Changes in the perforate behavior, in turn, affect the overall performance of the silencer. The focus of the present work is mean flow tangential to the perforate face, generally referred to as grazing flow. Mean flow that passes through the perforations, referred to as throughflow, may also be present. Due to the large flow losses incurred when forcing flow through the perforations, grazing flow is usually predominant, though some throughflow is inevitable and may be important (see, for example, Ref. 3). Of additional note, the high sound-pressure levels (SPLs) encountered are expected to cause nonlinear acoustic behavior for at least some of the perforations.

Although some degree of nonlinear perforate behavior can be expected in practice, linear approaches are common and useful tools in the design of automotive silencers. For a linear approach, the acoustic behavior of the perforations is represented in terms of a perforate impedance, given by

$$Z_p^{[f]} = \frac{\Delta p^{[f]}}{U_p^{[f]}} = i\chi^{[f]} + R^{[f]}, \tag{1}$$

where Δp is the acoustic pressure difference across the perforate, U_p is the acoustic velocity through the perforate, and i is the imaginary unit. The bracketed superscript is used here to denote a harmonic component at frequency f. As indicated in Eq. (1), the imaginary part of the impedance is defined as the reactance ($\chi^{[f]}$), and the real part of the impedance is defined as the resistance ($R^{[f]}$). Since the current study considers linear perforate behavior, frequency-domain analysis is implied and, for simplicity, the bracketed superscripts will be omitted hereafter.

In comparison to the limiting case of zero mean flow, the presence of significant grazing flow decreases the reactance and increases the resistance. The mechanisms behind these changes in impedance have been interpreted in several different ways (see, for example, the overview of works given by Cummings[4]). Often, the effects of grazing flow are simply described as a ''blowing away'' of the fluid in the orifice near field and its associated kinetic energy.[5,6] The decrease in χ and increase in R have been investigated in terms of the simultaneous deflections of the merging orifice and grazing fluid flows.[7,8] Other workers have related the grazing flow effects to the acoustically induced shedding of vorticity.[9,10] While the analytical methods used in the foregoing studies typically show qualitative agreement with experimental data,

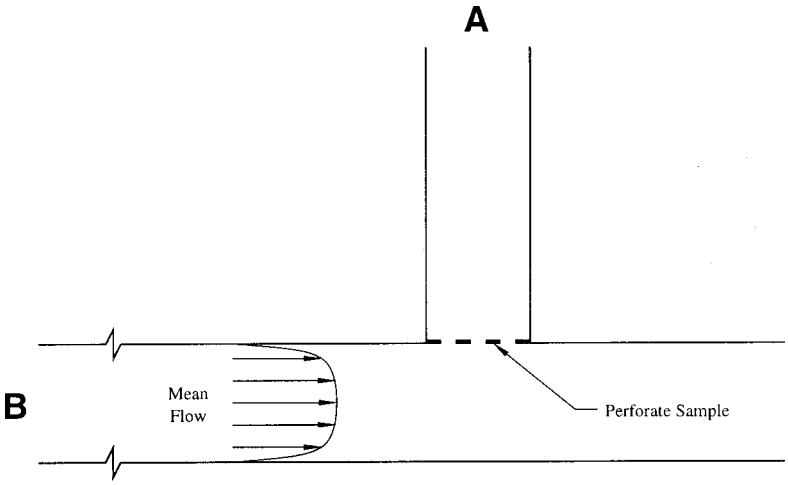


FIG. 1. Schematic of branch-type experimental setup.

the accuracy is questionable for the purpose of modeling perforated tube silencers. Thus, predictive tools must rely on experimental data and empirical models.

A number of workers have experimentally investigated the changes in impedance caused by grazing flow on one side of an orifice. Unfortunately, few of these studies consider orifice geometries and flow conditions representative of those in automotive silencers. Considering the differences in geometry and flow conditions of these experiments, it is not surprising that the resulting empirical correlations for perforate impedance often differ substantially. For example, several works use a mean velocity or Mach number to characterize the grazing flow.[11–13] However, for the case of interest here (perforate dimensions significantly smaller than the grazing flow boundary layer thickness), it has been shown that inner boundary layer parameters need to be included.[14,15] The friction velocity (u_τ) has been used to characterize the grazing flow and collapse impedance measurements reasonably well for a given perforate geometry.[4,15,16] However, the results of Kirby and Cummings[16] indicate that different empirical relationships are likely even for geometrically similar perforations of varying size.

The objective of the present investigation is to experimentally determine the impedance parameters of perforated samples exposed to grazing flow. A substantial portion of the study involves the design of an experimental apparatus for measuring the acoustic impedance of perforates exposed to flow conditions representative of those in the ducts of automotive systems. The results are presented in terms of nondimensional variables that collapse the data suitably for use in empirical perforate models.

II. EXPERIMENTAL APPROACH AND APPARATUS

In order to determine the impedance, three quantities are needed. The first two are the acoustic pressures on either side of the perforate sample, and the third is the acoustic velocity through the perforations. An experimental setup to determine these quantities was required which had good accuracy, allowed grazing flow and, in order to use an existing flow facility, used air as the working fluid. In addition, it was desirable to use an approach suitable for including throughflow in future experiments.

For measurements with grazing flow, a branch-type setup is attractive. In this setup, a main duct carries the mean flow and is separated from a branch duct (or cavity) by the perforated interface (see Fig. 1). The acoustic pressure on either side of the perforate can be measured directly. The perforate velocity, however, is determined by using a phase-pressure approach since direct measurement is difficult for realistic perforate geometries. In the phase-pressure approach, a linear acoustic analysis is performed for the branch duct. From this analysis, and known values of acoustic pressure or velocity at two locations in the branch duct (either through measurements or boundary conditions), the perforate velocity can be determined. Although axisymmetric cylindrical cavities have also been used,[17,18] the acoustic analysis is most easily performed for a branch duct with constant cross-sectional area.

There are three relatively common setups that have been used to apply the phase-pressure approach to the branch duct system. The first of these places the acoustic source (typically a loudspeaker) at the end of the branch (location A in Fig. 1) and uses a traversing probe microphone to measure acoustic pressures along the branch duct.[13,19,20] A second approach uses an oscillating rigid piston driven by an electromagnetic shaker for the acoustic source at location A.[11,14] Acoustic pressure from a fixed microphone in the branch duct combined with the measured piston velocity provide the two values needed in the analysis. The third common approach is to connect the acoustic source to the flow duct (location B in Fig. 1), and use a rigid closed end for the branch duct at location A.[4,18,21,22] Acoustic pressure measurements from a fixed microphone in the branch duct and the zero velocity boundary condition at the closed end provide the necessary information.

Each of the foregoing approaches has been applied effectively, but has certain advantages and disadvantages. The probe-tube technique is well-established and capable of good accuracy. However, the presence of the microphone and its traversing mechanism in the duct is less than ideal, particularly if throughflow is to be included. The oscillating rigid piston provides a clearly defined velocity boundary condition but is a much more complicated acoustic source to incorporate than a simple loudspeaker. Acoustic analysis of the sys-

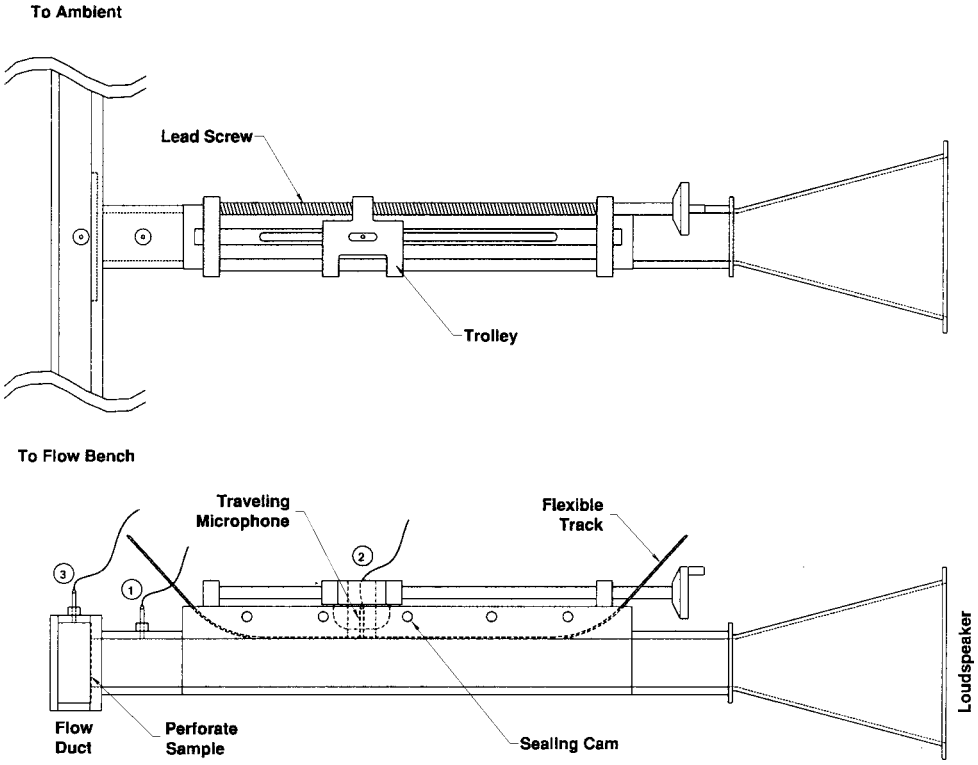


FIG. 2. Schematics of experimental setup.

tem with an oscillating piston boundary also becomes problematic if throughflow is considered. Rigid termination of the branch duct is simple, and very flexible if the cavity has an adjustable length so that its acoustic response may be varied. One potential drawback of the closed branch is that the acoustic waves from the source are propagating parallel to the perforate sample, and will therefore have phase and amplitude gradients in this direction. These gradients are inconsequential for studies of single perforations or very small samples, but can introduce errors for representative perforate samples. As with the oscillating piston, the acoustic analysis of the closed-end cavity becomes difficult if a source for throughflow is included.

For the present study, it was decided to place the acoustic source at the end of the branch duct and measure acoustic pressure at two locations in the branch duct. This approach is equivalent to using the two-microphone technique commonly used in impedance tube measurements.[23,24] A primary benefit of this configuration is that throughflow may be incorporated with less difficulty than the methods described previously. An error analysis of the system showed that axial location of the branch duct microphones can have a large effect on the accuracy of the measurements. Further investigation indicated that the effects of experimental errors could be minimized by locating one of the microphones near a pressure antinode for the standing waves in the branch duct. Similar findings have been reported with respect to traversing probe tube measurements.[25] The experimental apparatus was therefore designed with one microphone fixed, while the axial position of the other microphone could be varied.

A. Experimental apparatus

Figure 2 includes schematics of the experimental system design. The sidebranch duct is 7.62 cm square, and is centered in the longer wall of the rectangular flow duct having cross-sectional dimensions of 5.08 by 12.7 cm. The acoustic source (loudspeaker) is incorporated at the end of the branch duct opposite the perforate sample. The upstream side of the flow duct is 2.83 m in length to allow the turbulent boundary layer to develop. The downstream side of the flow duct is connected to a flow bench. Each perforate sample is bonded to a special fixture that is machined to ensure a flush interface between the sample and the wall of the flow duct. The sample fixture is attached to the wall of the flow duct, and all mating surfaces are sealed with a thin layer of vacuum grease.

Acoustic pressure measurements are made with three flush-mounted 0.64-cm condenser microphones as shown in Fig. 2. The axial position of microphone 1 is approximately 7.2 cm from the perforate sample, this value varying slightly with the thickness of the sample. Microphone 2 is installed in a fixture that attaches to a flexible ''T'' track such that the microphone face is flush with the bottom of the track. The flexible track slides in a groove milled in a solid piece of aluminum that forms the upper wall of the branch duct. A second piece of aluminum, milled to match the curved path of the track, applies pressure to the top of the flexible track through the action of several cams. Tightening the cams brings the base of the track flush with the inner surface of the branch duct, and seals the interface between the track and groove. The fixture that holds microphone 2 is inserted into a trolley that is connected to a precision lead screw. Releasing the cams and rotating the lead screw allows the axial position of microphone 2 to be adjusted (total travel is 36.8 cm). The third microphone is flush-mounted in the side of the flow duct.

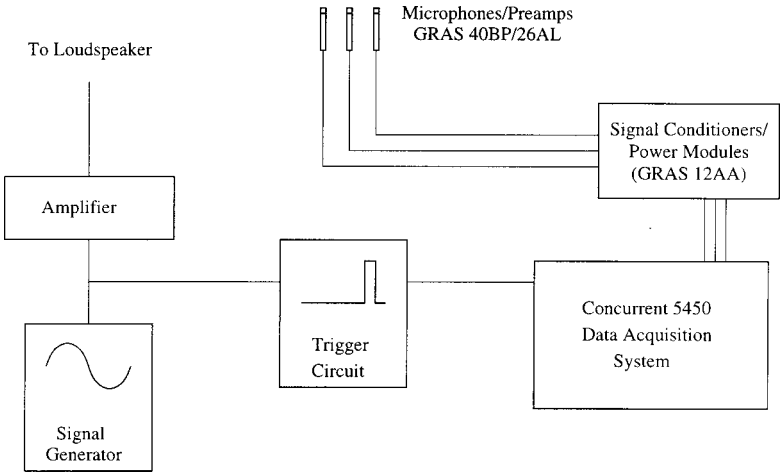


FIG. 3. Data acquisition setup for impedance measurements.

B. Impedance calculation

The sidebranch of the apparatus is analyzed using one-dimensional linear acoustic theory. The acoustic pressure and velocity at a location are given by

$$p = p_+ e^{-ikx} + p_- e^{ikx}, \tag{2}$$

and

$$u = \frac{1}{Y}(p_+ e^{-ikx} - p_- e^{ikx}), \tag{3}$$

respectively, where p_+ and p_- are the complex amplitudes of the positive and negative waves, respectively, k is the wave number, x is the axial coordinate, and Y is the characteristic impedance of the fluid (air). For the ideal case with no dissipation, $k = k_o = \omega/c_o$ and $Y = \rho_o c_o$ where $\omega = 2\pi f$ is the angular frequency, ρ_o is the mean density, and c_o is the mean speed of sound. Dissipation in the fluid can be considered by using a complex wave number and characteristic impedance as

$$k = k_o + \alpha - i\alpha, \quad Y = \rho_o c_o \left(1 - \frac{\alpha}{k_o} + i\frac{\alpha}{k_o}\right), \tag{4}$$

where α is the attenuation constant. For low frequencies, the attenuation constant may be approximated by[26]

$$\alpha \cong \frac{\sqrt{2\nu\omega}}{c_o d}, \tag{5}$$

where d is the pipe diameter and $\nu = \mu/\rho_o$ is the kinematic viscosity, μ being the absolute viscosity.

With the back surface of the perforate designated as $x = 0$, the two-microphone technique is used to determine p_+ and p_- as

$$p_+ = \frac{p_1}{e^{-ikl} + \mathcal{R}e^{ikl}}, \tag{6}$$

$$p_- = \frac{\mathcal{R}p_1}{e^{-ikl} + \mathcal{R}e^{ikl}}, \tag{7}$$

where p_n designates the measured acoustic pressure at microphone n, l is the distance from the back of the sample to microphone 1, and $\mathcal{R} = p_-/p_+$ is the reflection coefficient. The reflection coefficient is given by

$$\mathcal{R} = e^{-2ikl}\left(\frac{e^{-ik\Delta x} - \dfrac{p_2}{p_1}}{\dfrac{p_2}{p_1} - e^{ik\Delta x}}\right), \tag{8}$$

where Δx is the distance between microphones 1 and 2.

The perforate velocity is computed from the conservation of flow velocity between the perforate flow and 1D branch duct flow as

$$U_p A_p = u(x=0)A_d = \frac{A_d}{Y}(p_+ - p_-), \tag{9}$$

where A_p is the total perforate open area and A_d is the cross-sectional area of the branch duct. In terms of the sample porosity ($\sigma = A_p/A_d$), the perforate velocity is given by

$$U_p = \frac{1}{Y\sigma}(p_+ - p_-). \tag{10}$$

Finally, the perforate impedance is determined from Eq. (1), as

$$Z_p = \frac{\Delta p}{U_p} = \frac{p_3 - p(x = x_o)}{U_p}. \tag{11}$$

Inserting Eqs. (2), (6), (7), and (10) into Eq. (11) yields (after some manipulation)

$$\frac{Z_p}{Y\sigma} = i\frac{\dfrac{p_3}{p_1}\sin(k\Delta x) + \dfrac{p_2}{p_1}\sin(kl) - \sin(k(l+\Delta x))}{\cos(k(l+\Delta x)) - \dfrac{p_2}{p_1}\cos(kl)}. \tag{12}$$

C. Data acquisition

Figure 3 depicts the data acquisition setup for the measurements. Output from the three condenser microphones (GRAS 40BP with GRAS 26AL preamps) are amplified by GRAS 12AA microphone power modules. The signals from the amplifiers are sampled and stored by a Concurrent 5450 high-speed data acquisition system. In order to reduce the effects of cycle-to-cycle variations, data are taken for a number of cycles and phase averaged. The phase-locked averag-

ing process is made possible by a trigger circuit that initiates data acquisition at the same phase point for each cycle that is sampled.

The single frequency source signal is used for the input to a trigger circuit composed of a comparator with an adjustable voltage reference. The output from the trigger circuit is a rectangular wave that is at constant phase in relation to the source signal. A computer program, written specifically for data acquisition of cyclic processes (automotive intake/exhaust systems and silencers) controls the data acquisition system.[26] When the program is initiated, the first trigger pulse received begins the acquisition of data. Simultaneous data for three channels are acquired for a time interval slightly longer than one period of the source frequency. The trigger pulse at the end of a data acquisition cycle is ignored and the subsequent pulse starts the process over. Data for 512 cycles are obtained and the 512 measurements at each phase point are averaged. A second computer program is then used to convert the voltage measurements to pressure values.

For simultaneous measurement of three channels, the data acquisition system can sample at rates up to 167 kHz. However, the phase-averaging method reduced the actual sampling rate to roughly 100 kHz, depending somewhat on the source frequency. At the highest source frequency of the study (975 Hz), a minimum of 96 data points per acoustic cycle was obtained, ensuring that waveform resolution was not an issue.

D. Experimental procedure

Agreement between the microphones was tested by installing them in a flat-faced end cap that was mounted perpendicular to the branch duct axis. Calibration measurements were made with pure-tone signals for frequencies between 100 and 1000 Hz. The microphones were calibrated for amplitude and phase within 1% and 0.5 deg, respectively.

To perform a test, the flow bench was first set at the grazing flow velocity of interest. After setting the signal generator to the desired frequency, the trigger circuit was adjusted to provide a narrow rectangular wave output. The axial position of microphone 2 was moved to a pressure minima location, and the data acquisition program was initiated. When the sampling was complete, the acquisition program performed the phase-averaging process and stored both the raw data for each cycle and the averaged data to an internal disk. At each frequency/grazing flow combination, at least three measurements were taken with the position of microphone 2 varied slightly. This procedure allows verification that the pressure minima was located, and also provides information regarding repeatability of the measurement.

The harmonic components of the measured pressures were obtained by postprocessing the data with a fast Fourier transform (FFT) algorithm. In order to apply the FFT with a rectangular window, 2^N points are required for a single period of the fundamental frequency, where N is the integer order of the FFT. Linear interpolation was used to construct the required number of points from the measured data. Since the experimental data were smooth and free of discontinuities, the linear interpolation scheme is adequate. This fact was confirmed by reconverting the FFT output to the time domain and comparing with the original pressure traces.

TABLE I. Dimensions of the perforate samples (lengths in cm).

Sample	Effective diameter (d_p)	Wall thickness (t_w)	Porosity (σ)
Orifice	0.249	0.091	0.0210
Louver 1	0.297	0.1	0.0383
Louver 2	0.225	0.1	0.0218

During initial experiments, it was found that reflections from the ends of the main flow duct can affect the accuracy of the measurements. As mentioned in the discussion of experimental setups, waves propagating in the main flow duct are traveling parallel to the sample, rather than normal to it. Ideally, the wave amplitudes in the flow duct would be minimized, and the acoustic pressure at microphone 3 will be much smaller than within the branch duct. At certain frequencies, however, reflections from the ends of the main flow duct cause the acoustic pressure at microphone 3 to be nearly equal to that in the branch duct. To minimize the effects of these reflections, measurement frequencies were selected such that a pressure minima for the flow duct was located at microphone 3, a condition that occurred at roughly 60-Hz intervals. This condition did not diminish the frequency resolution of the measurements substantially, but the exact frequency for a measurement could not be selected arbitrarily.

III. RESULTS AND DISCUSSION

Measurements are presented for three perforate samples. The first sample is a circular, square-edged orifice, while the other two are stamped louver geometries used in production automotive mufflers. Geometric parameters for the samples are included in Table I. The orifice dimensions are reasonably representative of perforations in automotive silencers, and are very similar to the orifice geometry used in a number of silencer investigations.[17,27,28] The orifices were placed in staggered rows such that the minimum distance between orifice centers was approximately five orifice diameters. The basic layout of both louver samples is similar (see Fig. 4), but the individual perforations (openings) for the two louver samples have distinctly different geometries. Both louver samples consisted of four rows of louvers aligned with the flow direction. Each row of louvers contained four pairs of perforations (eight openings total), with each pair opening in a ''butterfly'' configuration from the centerline of the row. The louvers were oriented such that the stamped material projected into the branch duct. An effective diameter for a single louver opening (d_p) is computed as

$$d_p = \sqrt{\frac{4A_{o,p}}{\pi}}, \tag{13}$$

where $A_{o,p}$ is the cross-sectional area of a single opening. Wall thickness for the louvers represents a nominal material thickness. Due to the angled geometry of the louvers, a true ''flow path length'' is somewhat longer than the material thickness, but is difficult to specify.

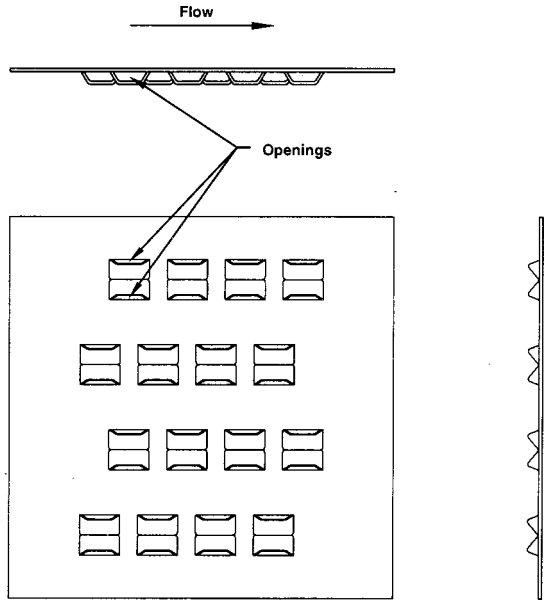


FIG. 4. Schematic showing layout of the louver samples.

The experimental data include measurements for frequencies between 200 and 1000 Hz and grazing flow velocities between 0 and 36.6 m/s. For presentation, the data are reduced in terms of an equivalent length (l_{eq}) and resistance (R) for a single opening. The equivalent length is determined from the reactance as

$$l_{\mathrm{eq}}=\frac{\chi}{\rho_o\omega}, \tag{14}$$

and the resistance is simply the real part of the impedance as given by Eq. (1). For linear behavior of an isolated circular orifice with zero grazing flow, an approximation for l_{eq} can be obtained from the Rayleigh end correction as

$$l_{\mathrm{eq}}\cong t_w+0.85d_p, \tag{15}$$

and R can be estimated from the semiempirical expression given by

$$R\cong\rho_o\sqrt{\nu\omega}\sqrt{8}\left(1+\frac{t_w}{d_p}\right). \tag{16}$$

These expressions are based on a number of assumptions as discussed by Melling,[29] but provide representative values for the impedance parameters with no grazing flow. As mentioned earlier, in the presence of significant grazing flow l_{eq} decreases and R increases from the zero flow values.

The measured impedance data are nondimensionalized and, for all three perforates, are collapsed using a dimensionless grazing flow velocity parameter. The scaling parameters are similar to those used by Kirby and Cummings,[16] and use the friction velocity to characterize the grazing duct flow. The friction velocity for the duct flow is computed as

$$u_\tau=\sqrt{\frac{\tau_w}{\rho_o}}, \tag{17}$$

where τ_w is the wall shear stress. The wall shear stress is determined from the mean grazing flow velocity (U_g) and the Darcy friction factor (f_D) as

$$\tau_w=\rho_o U_g^2\frac{f_D}{8}. \tag{18}$$

The value of f_D for a flow in a rectangular duct is computed using the expression

$$f_D=\frac{0.178}{\mathrm{Re}_d^{1/5}}, \tag{19}$$

where Re_d is the Reynolds number. Note that for the rectangular duct, the constant in Eq. (19) is reduced slightly from the common expression used for a circular duct (see, for example, Refs. 4, 30, 31). For the experiments presented in this study, $0\leqslant u_\tau\leqslant 1.6$. In comparison to conditions in an automotive exhaust muffler, these values for u_τ are representative of lower operating speeds and loads. Based on the experimental data of Selamet *et al.*,[32] $1\leqslant u_\tau\leqslant 8$ is an approximate range for a 3.0L engine running at wide-open throttle between 1000 and 5000 revolutions per minute. For part load conditions, the values of u_τ will be lower.

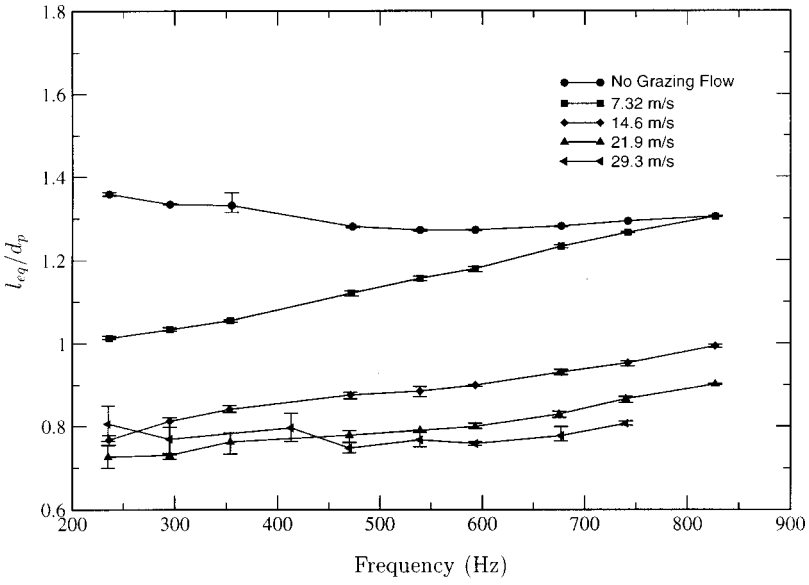


FIG. 5. Nondimensionalized equivalent length with varied grazing flow velocity plotted against frequency (circular orifice).

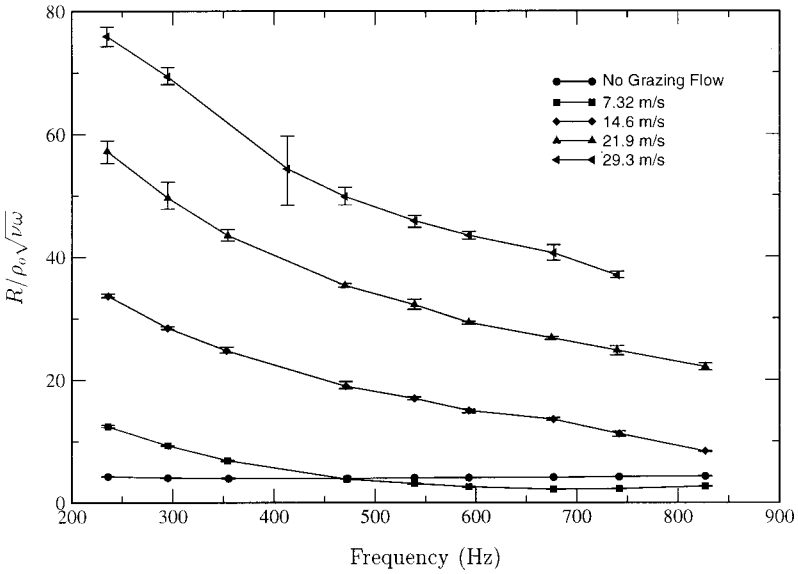


FIG. 6. Nondimensionalized resistance with varied grazing flow velocity plotted against frequency (circular orifice).

A. Circular orifice

To demonstrate the changes caused by grazing flow, Figs. 5 and 6 include the circular orifice measurements plotted in a "raw" form that does not include the friction velocity. The equivalent length is nondimensionalized with d_p and the resistance is nondimensionalized with $\rho_o\sqrt{\nu\omega}$. For each frequency/grazing flow combination, the values for l_{eq} and R are computed as the average of all measurements taken (between three and five values). Error bars in the data plots indicate the range of values obtained.

With the selection of axes in Figs. 5 and 6, clear trends are rather difficult to distinguish. However, the general effects of grazing flow on l_{eq} and R are evident. For the zero-flow case, l_{eq} is essentially constant, and is then reduced by the grazing flow. At a particular grazing flow velocity, the reduction in l_{eq} is larger for lower frequencies. $R/\rho_o\sqrt{\nu\omega}$ is also nearly constant for the zero-flow case. At the lower flow rates, $R/\rho_o\sqrt{\nu\omega}$ decreases slightly, but a point is reached where it consistently increases with increasing grazing flow velocity. Similar to the l_{eq} measurements, the changes caused by a specific grazing flow velocity become larger as frequency decreases.

In Figs. 7 and 8, the impedance data of Figs. 5 and 6 have been parametrized to account for the grazing flow. The equivalent length is again nondimensionalized by d_p while the resistance is scaled by $\rho_o\omega d_p$. Both parameters are plotted against $u_\tau/\omega d_p$, which collapses the data reasonably well, although some deviations are evident. Since the orifice is a commonly used geometry in the study of perforated tube silencers, empirical expressions to fit the data have been formulated and are included in the figures.

The equivalent length remains relatively constant for values of $u_\tau/\omega d_p$ below approximately 0.03. Above this point, a rapid decrease is seen as the effect of grazing flow becomes significant. At higher values of $u_\tau/\omega d_p$, the equivalent length appears to reach a lower limit reduced from the zero-flow case by about 40%. This reduction is slightly greater than the classic Rayleigh end correction for

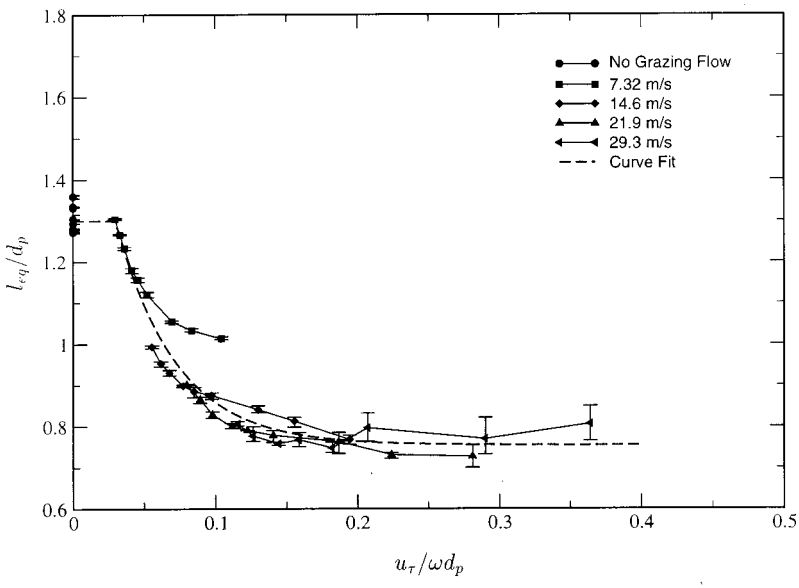


FIG. 7. Nondimensionalized equivalent length with varied grazing flow velocity plotted against nondimensionalized friction velocity (circular orifice).

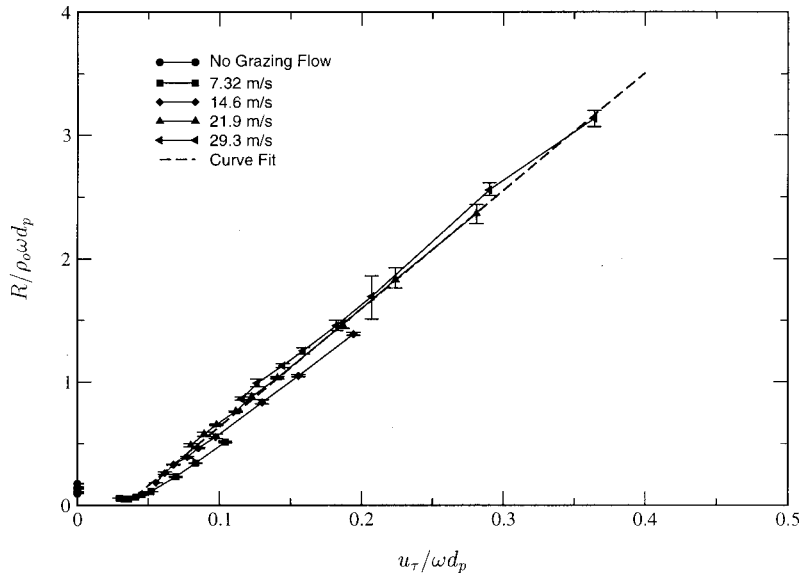


FIG. 8. Nondimensionalized resistance with varied grazing flow velocity plotted against nondimensionalized friction velocity (circular orifice).

one side of the orifice ($0.425\ d_p$, or approximately 37% of l_{eq}/d_p for this orifice). In the approximate range $0.06 \leqslant u_\tau/\omega d_p \leqslant 0.1$, the measurements at $U_g = 7.32$ m/s are noticeably higher than the other results. Specific reasons for this are uncertain, but it is not surprising that this rather simple parametrization does not fully describe the physics. The curve fit shown in Fig. 7 is given by

$$\frac{l_{\mathrm{eq}}}{l_{\mathrm{eq},0}} = \begin{cases} 1.0 & 0 \leqslant u_\tau/\omega d_p \leqslant 0.03 \\ 0.58 + 0.42e^{-23.6(u_\tau/\omega d_p - 0.03)} & 0.03 < u_\tau/\omega d_p \end{cases}, \quad (20)$$

where $l_{\mathrm{eq},0}$ is the zero-flow equivalent length.

For the resistance measurements, $R/\rho_o\omega d_p$ first appears to decrease by a small amount between the zero-flow limit and $u_\tau/\omega d_p \cong 0.04$. In this range such changes are likely to be rather inconsequential, since the reactive part of the impedance is expected to dominate. For larger values of $u_\tau/\omega d_p$, the grazing flow effects become more significant. The value of $R/\rho_o\omega d_p$ increases in a nearly linear fashion with $u_\tau/\omega d_p$ and, for the highest value of $u_\tau/\omega d_p$, is increased from the zero-flow limit by more than an order of magnitude. Relatively small but consistent differences in the results for $u_\tau/\omega d_p$ less than approximately 0.2 further suggest an incomplete characterization of the physics. These differences appear to become negligible at the higher grazing flow velocities. For $u_\tau/\omega d_p \geqslant 0.05$, a linear curve fit to the data yields

$$\frac{R}{\rho\omega d_p} = -0.32 + 9.57\frac{u_\tau}{\omega d_p}, \quad (21)$$

which is included in Fig. 8.

B. Louver 1

Figures 9 and 10 include the scaled impedance parameters for the first louver sample. For the equivalent length, small variations occur for values of $u_\tau/\omega d_p$ below approximately 0.08. The values then decrease rapidly up to approximately $u_\tau/\omega d_p = 0.16$. Above this point, l_{eq} appears to reach a lower limit reduced from the zero-flow case by roughly $0.35 d_p$, an amount slightly less than the Rayleigh end correction for one side of an orifice of the same diameter. While

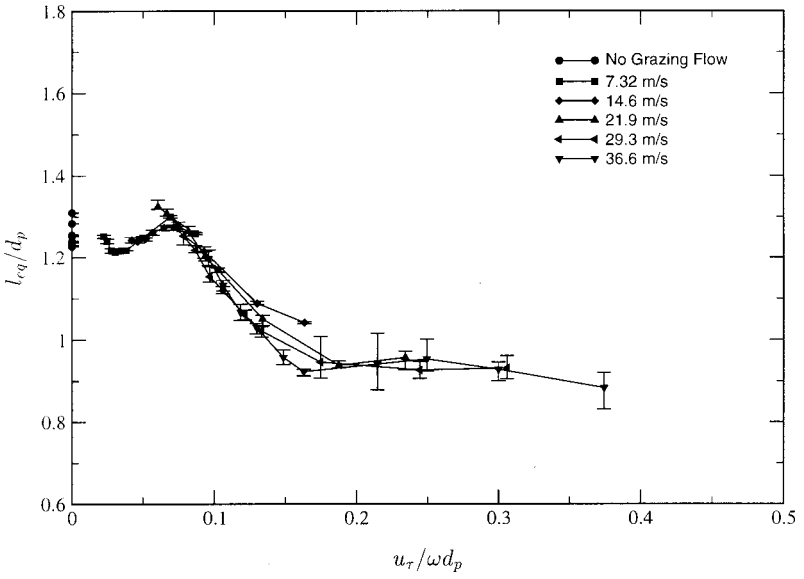


FIG. 9. Nondimensionalized equivalent length with varied grazing flow velocity plotted against nondimensionalized friction velocity (louver 1).

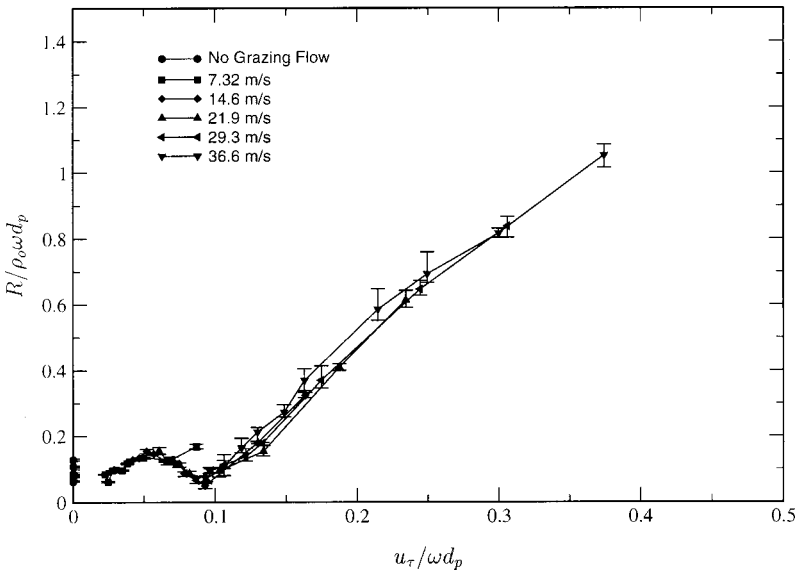


FIG. 10. Nondimensionalized resistance with varied grazing flow velocity plotted against nondimensionalized friction velocity (louver 1).

the basic trends of the louver are similar to the circular orifice, details in the behavior are noticeably different. In particular, the value of $u_\tau/\omega d_p$ where l_{eq}/d_p begins to decrease is significantly higher, while the point at which the lower limit is reached is rather close to that for the orifice.

In the resistance values for louver 1, a small increase and decrease in $R/\rho_o\omega d_p$ is seen between the zero flow limit and $u_\tau/\omega d_p \cong 0.1$. While this behavior is rather interesting, it is not expected to have much influence on the behavior of a perforated tube silencer. For larger values of $u_\tau/\omega d_p$, the value of $R/\rho_o\omega d_p$ increases rapidly and reaches a point where it increases quite linearly with $u_\tau/\omega d_p$. The amount of increase in $R/\rho_o\omega d_p$ indicates that, with respect to the resistance, this louver is less sensitive to grazing flow than the circular orifice.

C. Louver 2

Figures 11 and 12 include the scaled impedance parameters for louver 2. The reduction in l_{eq}/d_p for this perforate is not as rapid or well-defined as for the other two samples. The data also appear somewhat more erratic, but for a given value of $u_\tau/\omega d_p$, most points are within a $\pm 5\%$ range. It is not clear whether a lower limit for l_{eq}/d_p has been reached at the highest value of $u_\tau/\omega d_p$. The measured behaviors of the other samples indicate that a lower limit is likely, and the amount of reduction in l_{eq} suggests that a lower limit was nearly, if not fully, reached. However, additional measurements at higher values of $u_\tau/\omega d_p$ would be necessary to confirm this supposition.

The behavior of the resistance for louver 2 is also noticeably different from the other two samples. Except for an initial slight decrease at low flow rates, the resistance increases steadily with increasing $u_\tau/\omega d_p$. For the range of data obtained, however, it appears that the increase in $R/\rho_o\omega d_p$ is not quite linear with $u_\tau/\omega d_p$, but to a power slightly greater than 1. This behavior is consistent with the louver data presented by Kirby and Cummings.[16] It is uncertain whether a more linear relationship would be evident with higher grazing flow velocities.

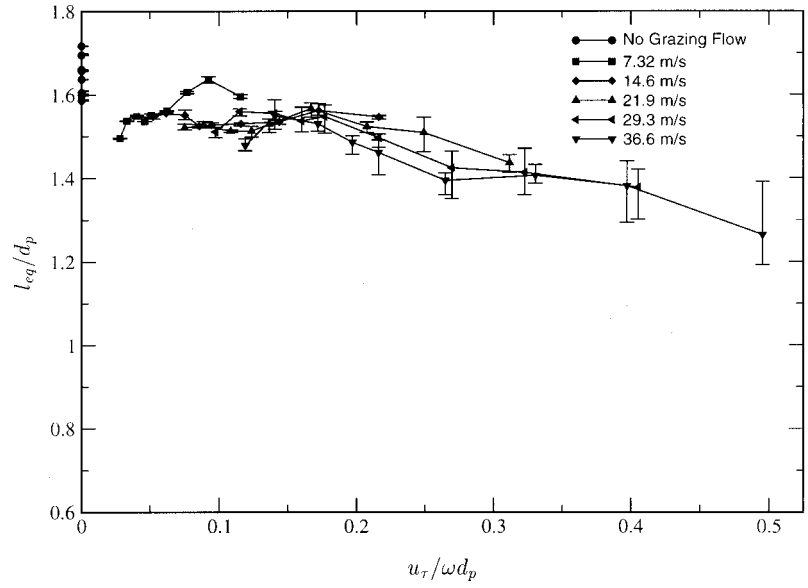


FIG. 11. Nondimensionalized equivalent length with varied grazing flow velocity plotted against nondimensionalized friction velocity (louver 2).

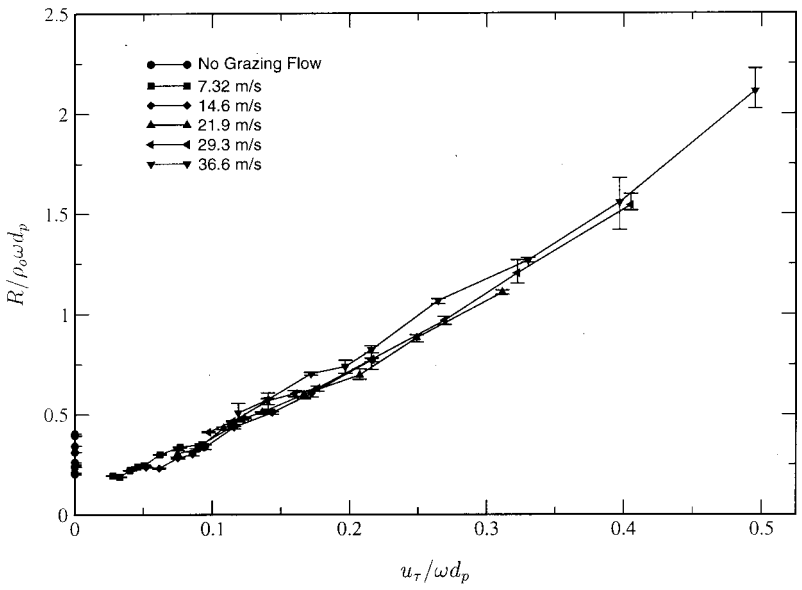


FIG. 12. Nondimensionalized resistance with varied grazing flow velocity plotted against nondimensionalized friction velocity (louver 2).

IV. CONCLUDING REMARKS

A branch-type system has been designed to measure the impedance of perforated plates exposed to grazing flow. In the branch duct, the system uses one fixed and one movable microphone. Locating the movable microphone near an antinode for the acoustic pressure improves the error characteristics of the system. The system was used successfully to measure the impedance of perforated plates exposed to grazing flow. An additional feature of the setup, to be investigated in future studies, is that throughflow may be included with less difficulty than other branch-type systems.

The experimental system was used to obtain the impedance parameters of three perforate samples. The data have been scaled using the friction velocity to characterize the grazing flow. At higher $u_\tau/\omega d_p$ the equivalent length of each sample decreased by roughly $0.3-0.4d_p$, amounts fairly representative of the Rayleigh end correction for one side of an orifice. Resistance of the louvers was somewhat less sensitive to grazing flow than the circular orifice, though all three samples showed significant increases in resistance with grazing flow. Consistent with the findings of Kirby and Cummings,[16] behaviors of the different samples showed distinct differences, indicating that using a single set of empirical expressions for all perforates is a questionable approach. Measurements of perforate impedance with grazing flow are useful for studying the effects of grazing flow on silencer performance. However, the differences in results obtained for different perforations suggest that in order to most accurately model a specific silencer, experimental data should be obtained for the particular perforations used in that silencer.

As might be expected, relatively small but systematic differences in the scaled results indicate that the simple parametrization used does not fully characterize the physics. Nevertheless, the equivalent length and resistance collapse reasonably well with the scaling factors used. Although the conditions in an automotive muffler differ from the experimental conditions in this study, the nondimensional parameters of interest are the same. Therefore, provided that basic assumptions of this study are met (no throughflow and predominantly linear behavior), the data are expected to provide impedance parameters that are suitable for use in engineering calculations.

[1] M. L. Munjal, "Analysis of a flush-tube three-pass perforated element muffler by means of transfer matrices," Int. J. Acoust Vib. **2**, 63–68 (1997).
[2] N. S. Dickey, A. Selamet, and J. M. Novak, "Multi-pass perforated tube silencers: A computational approach," J. Sound Vib. **211**, 435–448 (1998).
[3] J. W. Sullivan, "Some gas flow and acoustic pressure measurements inside a concentric-tube resonator," J. Acoust. Soc. Am. **76**, 479–484 (1984).
[4] A. Cummings, "The effects of grazing turbulent pipe flow on the impedance of an orifice," Acustica **61**, 233–242 (1986).
[5] D. Ronneberger, "The acoustical impedance of holes in the wall of flow ducts," J. Sound Vib. **24**, 133–150 (1972).
[6] Y. Hirata and T. Itow, "Influence of air flow on the attenuation characteristics of resonator type mufflers," Acustica **28**, 115–120 (1973).
[7] B. E. Walker and A. F. Charwat, "Correlation of the effects of grazing flow on the impedance of Helmholtz resonators," J. Acoust. Soc. Am. **72**, 550–555 (1982).
[8] T. Rogers and A. S. Hersch, "Effect of grazing flow on the steady-state resistance of square-edged orifices," in AIAA 2nd Aero-Acoustics Conference, 1975, pp. 259–287.
[9] M. S. Howe, "The influence of grazing flow on the acoustic impedance of a cylindrical wall cavity," J. Sound Vib. **67**, 533–544 (1979).
[10] P. Mungur and J. L. Whitesides, "Influence of grazing flow on duct wall normal impedances," in AIAA 2nd Aero-Acoustics Conference, 1975, pp. 289–306.
[11] B. Phillips, "Effects of high-wave amplitude and mean flow on a Helmholtz resonator," NASA TM X-1582, 1968.
[12] G. D. Garrison, A. C. Schnell, C. D. Baldwin, and P. R. Russell, "Suppression of combustion oscillations with mechanical damping devices," Pratt and Whitney Report PWA FR-3299, 1969.
[13] K. N. Rao and M. L. Munjal, "Experimental evaluation of impedance of perforates with grazing flow," J. Sound Vib. **108**, 283–295 (1986).
[14] A. L. Goldman and R. L. Panton, "Measurement of the acoustic impedance of an orifice under a turbulent boundary layer," J. Acoust. Soc. Am. **60**, 1397–1404 (1976).
[15] A. L. Goldman and C. H. Chung, "Impedance of an orifice under a turbulent boundary layer with pressure gradient," J. Acoust. Soc. Am. **71**, 573–579 (1982).
[16] R. Kirby and A. Cummings, "The impedance of perforated plates subjected to grazing gas flow and backed by porous media," J. Sound Vib. **217**, 619–636 (1998).
[17] J. W. Sullivan and M. J. Crocker, "Analysis of concentric-tube resonators having unpartitioned cavities," J. Acoust. Soc. Am. **64**, 207–215 (1978).

[18] N. S. Dickey, ''A Time Domain Approach for Acoustic Analysis of Perforated Tube Silencers,'' Ph.D. thesis, The University of Michigan, Ann Arbor, MI, 1999.
[19] P. A. Marino, N. Bohn, P. L. Russell, A. C. Schnell, and G. L. Parsons, ''A study of the suppression of combustion oscillations with mechanical damping devices,'' NASA CR-90094, 1967.
[20] E. Feder and L. W. Dean III, ''Analytical and experimental studies for predicting noise attenuation in acoustically treated ducts for turbofan engines,'' NASA CR-1373, 1969.
[21] J. W. Sullivan, ''Theory and methods for modelling acoustically long, unpartitioned cavity resonators for engine exhaust systems,'' Ph.D. thesis, Purdue University, West Lafayette, IN, 1974.
[22] P. D. Dean, ''An *in situ* method of wall acoustic impedance measurement in flow ducts,'' J. Sound Vib. **34**, 97–130 (1974).
[23] A. F. Seybert and D. F. Ross, ''Experimental determination of acoustic properties using a two-microphone random-excitation technique,'' J. Acoust. Soc. Am. **61**, 1362–1370 (1977).
[24] J. Y. Chung and D. A. Blaser, ''Transfer function method of measuring in-duct acoustic properties. I. Theory, pp. 907–913. II. Experiment, pp. 914–921'' J. Acoust. Soc. Am. **68**, 907–921 (1980).
[25] M. L. Munjal, *Acoustics of Ducts and Mufflers* (Wiley, New York, 1987).
[26] V. Kothamasu, ''Effect of intake and exhaust elements on sound attenuation and engine performance: An experimental and computational investigation,'' Master's thesis, The Ohio State University, Columbus, OH, 1999.
[27] J. W. Sullivan, ''A method for modeling perforated tube muffler components. I. Theory, pp. 772-778. II. Applications, pp. 779–788,'' J. Acoust. Soc. Am. **66**, 772–788 (1979).
[28] N. S. Dickey and A. Selamet, ''Acoustic nonlinearity of a circular orifice: An experimental study of the instantaneous pressure/flow relationship,'' Noise Control Eng. J. **46**, 97–107 (1998).
[29] T. H. Melling, ''The acoustic impendance of perforates at medium and high sound pressure levels,'' J. Sound Vib. **29**, 1–65 (1973).
[30] H. K. Myong and T. Kobayashi, ''Numerical simulation of three-dimensional developing turbulet flows in a square duct with the anisotropic κ–ϵ model,'' in *Advances in Numerical Simulation of Turbulent Flows [ASME/JSME Fluids Engineering Conference], Portland, OR, June 1991*, FED Vol. 117, pp. 17–23.
[31] H. Fujita, ''Turbulent flow in smooth and rough-walled square ducts,'' Trans. Jpn. Soc. Mech. Eng., Ser. B **45**, 197 (1979).
[32] A. Selamet, S. H. Yonak, J. M. Novak, and M. Khan, ''The effect of vehicle exhaust system components on flow losses and noise in firing spark-ignition engines,'' SAE 951260, 1995.

The acoustic properties of granular materials with pore size distribution close to log-normal

K. V. Horoshenkov[a] and M. J. Swift
School of Engineering, University of Bradford, Bradford BD7 1DP, England

(Received 3 April 2001; accepted for publication 9 August 2001)

The majority of realistic porous materials are composed of pores of which the shape is variable and the size of the pores normally obeys a distinctive statistical distribution. Although the variation of pore shape is less important, the statistical parameters of pore size distribution can have a considerable effect on the acoustic properties of porous media. This paper discusses the application of a simple model for the prediction of the acoustic properties of porous granular media with some assumed pore geometry and pore size distribution close to log-normal. The model is based on the rational (Padé) approximation approach [K. V. Horoshenkov, K. Attenborough, and S. N. Chandler-Wilde, J. Acoust. Soc. Am. **104**, 1198–1209 (1998)] which has been developed for some simple pore geometries. It is shown that the experimentally determined pore size distribution for a representative range of granular materials is often close to log-normal. This assumption enables accurate predictions of the acoustic performance of these materials using the presented model. The water suction method is proposed to determine the parameters of the log-normal distribution, which are the mean pore size, $\langle \phi \rangle$, and its standard deviation, σ. This method is nonacoustic, modelless and well-adapted to acoustic materials and, unlike the BET method [S. Brunauer, P. H. Emmett, and E. Teller, J. Am. Chem. Soc. **60**, 309–319 (1938)], is easy to reproduce in any basic acoustic laboratory requiring no expensive parts or chemicals. The proposed Padé approximation is based entirely on four measurable nonacoustic parameters, the porosity, Ω, flow resistivity, R_b, tortuosity, q^2 and the standard deviation of the pore size, σ. The method is successfully tested on a representative selection of consolidated and nonconsolidated porous granular materials. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1408312]

PACS numbers: 43.50.Gf, 43.20.Gp, 43.55.Ev [MRS]

I. INTRODUCTION

A standard approach for sound propagation in a rigid-frame porous material has been to model the fluid-filled medium as a homogeneous fluid with complex, frequency dependent characteristic impedance (Z_b) and propagation constant (k_b). Many available models adopt a theoretical microstructural approach to include viscous and thermal effects in tortuous pores of different geometries. These effects are normally treated separately by introducing complex expressions for the dynamic density and compressibility of the fluid. Expressions from these models are based upon the exact equations from the Kirchhoff theory and that of Zwikker and Kosten,[1] which assume that rigid-frame porous media can be modelled as a stack of parallel capillary tubes. This view that rigid-frame porous media can be modelled as a stack of parallel capillary tubes has been taken by many researchers.[1–5] There is a considerable amount of experimental evidence which confirms that this assumption is able to provide a good basis for developing reliable theoretical models for the prediction of the acoustic properties of porous surfaces.[4,6]

On the other hand, many theoretical models are more difficult to compute than the simple empirical models of Delany and Bazley[7] and Miki,[8] or the semiempirical models of Allard[9] and Wilson,[10] in that they involve a ratio of Bessel functions, $J_1(z)/J_0(z)$, for complex values of z, with argument $\pm \pi/4$. An excellent model proposed by Allard[11] relies upon a knowledge of the viscous and thermal characteristic lengths, which can be difficult to determine. Models by Attenborough,[4] Stinson,[6] and Wilson[10] involve empirical shape factors, which values are adjusted liberally in many cases to fit the experimental data. Another model has been presented by Attenborough[12] which accounts for the effect of the statistically distributed pore size and which may be considered more practical, but still complex to implement because it requires the numerical integration of exponential functions of complex argument. The importance of the statistical approach is dictated by the fact that in the real world the majority of porous materials are composed of pores of which the shape is variable and the size of the pores normally obeys a distinctive statistical distribution. Although it has been suggested that the shape of pores is relatively unimportant,[12] for these materials we should only refer to the values of the dynamic density and complex compressibility which have been averaged over a range of realistic pore sizes identified in the material. The same argument is applicable to the acoustic quantities such as characteristic impedance and propagation constant.

Various approximations of the theoretical microstructural models as well as the empirical models are often regarded as much more attractive in acoustical problems which involve large amounts of computation. It is obvious that

[a]Author to whom correspondence should be addressed. Electronic mail: k.horoshenkov@bradford.ac.uk

cumbersome expressions defining the boundary conditions in a realization of the finite and boundary element methods for outdoor sound propagation or expressions for the reconstruction of the surface impedance can seriously affect the computation time or even make the problem virtually impossible to solve. A rational (Padé) approximation has recently been proposed.[13] This approximation provides a simple and reliable method for predicting the acoustic properties of rigid-frame porous media with log-normally distributed pore size. Unlike some other models which include empirical shape factors the proposed Padé approximation is based entirely on a set of four measurable nonacoustic parameters, which are the porosity, Ω, flow resistivity, R_b, tortuosity, q^2 and the standard deviation of the pore size, σ. These parameters are determined routinely using a basic experimental setup.

This paper studies the application of the Padé approximation model to the prediction of the acoustic properties of a class of granular materials. In Sec. I the theoretical formulation for sound propagation in porous media with a statistically distributed pore size is revised. This formulation provides a more general explanation of the approach by Yamamoto.[14] Section II provides the expressions and tables with the coefficients for the Padé approximants[13] for three pore geometries. In Sec. III the details of the experimental procedure to determine the nonacoustic properties of porous materials are discussed. In Sec. IV the theoretical predictions for the acoustic properties of some porous granular materials are compared against the experimental results.

II. THEORETICAL FORMULATION

We consider that a rigid frame porous sample of cross-sectional area S_0 is pierced through by uniform pores of various cross-sectional areas and that the pores can be modelled as a stack of capillary tubes of particular shape which size obeys a distinctive statistical distribution. For any fluid property ψ we let $\overline{\psi}$ denote the mean value over the pore cross section and $\langle\psi\rangle$ denote the mean value of $\overline{\psi}$ over all pore sizes.

The fluid velocity in an individual pore $u_i(x,y,\omega)$ is found as the solution of the equation of motion

$$\rho_0 \frac{\partial u_i}{\partial t} + \mu \Delta u_i + \nabla p = 0. \tag{1}$$

Here ρ_0 is the equilibrium density of air, μ is the dynamic viscosity of air, p is the loading acoustic pressure, and ω is the circular frequency. The system of coordinates (x,y) is chosen so that its plane is parallel to the surface of the porous sample.

For every particular pore area s_i the average seepage velocity is found by integrating u_i within the limits defining the pore section. The average velocity of fluid through the ith pore is then predicted by integrating the equation of motion (1) over the cross-sectional area of this pore s_i so that

$$\frac{i\omega\rho_0}{s_i}\int\int_{s_i} u_i(x,y,\omega)dx\,dy - \frac{\mu}{s_i}\int\int_{s_i}\Delta u_i(x,y,\omega)dx\,dy = \nabla p. \tag{2}$$

Let us now assume that the cross-sectional area s_i of every individual pore is a unique value and can be statistically predicted with the probability

$$\chi_i = \frac{s_i}{S_p}, \tag{3}$$

where $S_p = \Sigma_i s_i$ is the total cross-sectional area of the sample of material occupied by the pores and the porosity of the material is $\Omega = S_p/S_0$. In this case the average seepage velocity of the fluid through the porous area of the sample S_p can be found from the continuity equation as

$$\langle u_a\rangle = \frac{1}{S_p}\sum_{i=1}^{N} s_i\overline{u_i} = \sum_{i=1}^{N}\chi_i\overline{u_i} \tag{4}$$

or

$$i\omega\rho_0\sum_{i=1}^{N}\frac{\chi_i}{s_i}\int\int_{s_i} u_i(x,y,\omega)dx\,dy - \mu\sum_{i=1}^{N}\frac{\chi_i}{s_i}\int\int_{s_i}\Delta u_i(x,y,\omega)dx\,dy = \nabla p, \tag{5}$$

where N is the number of pores within the area S_0.

From the integration and the Gauss integral theorem we obtain

$$i\omega\rho_0\sum_{i=1}^{N}\chi_i\overline{u_i}(\omega) - \mu\sum_{i=1}^{N}\frac{\chi_i}{s_i}\int_{\Gamma_i}\nabla u_i(x,y,\omega)dx\,dy = \nabla p, \tag{6}$$

where Γ_i is the circumference of the ith pore, which depends only upon the pore shape and can be explicitly expressed for a few particular pore geometries.

When $S_0\to\infty$, $N\to\infty$, then the sums should be replaced by integrals and the average velocity is weighted by the probability density function for the distribution $e(s)$ of the size of identically shaped pores in the material. Thus expression (6) is transformed into

$$i\omega\rho_0\int_o^\infty e(s)\overline{u}(s,\omega)ds - \mu\int_o^\infty s^{-1}e(s)\times\left[\int_\Gamma\nabla u(x,y,\omega)dx\,dy\right]ds = \nabla p. \tag{7}$$

The integral in the first term in (7),

$$\langle u_a(\omega)\rangle = \int_0^\infty e(s)\overline{u}(s,\omega)ds, \tag{8}$$

is the average seepage velocity per unit cross-sectional area and s is a characteristic dimension of the pore of the assumed geometry with pore circumference Γ. The second term

$$\langle\tau_a(\omega)\rangle = \mu\int_0^\infty s^{-1}e(s)\left[\int_\Gamma\nabla u(x,y,\omega)dx\,dy\right]ds \tag{9}$$

is the average viscous friction force per unit cross-sectional area. Following the work by Biot[3] it is possible to use expressions (8) and (9) to derive the viscosity correction func-

tion for sound propagation in a material with identically shaped pores but variable in size

$$i\omega\rho_0\left(1-\frac{1}{i\omega\rho_0}\frac{\langle\tau_a(\omega)\rangle}{\langle u_a(\omega)\rangle}\right)\langle u_a(\omega)\rangle=\nabla p. \tag{10}$$

In its limiting case the above expression reduces to

$$\lim_{\omega\to 0}\frac{\langle\tau_a(\omega)\rangle}{\langle u_a(\omega)\rangle}=\lim_{\omega\to 0}\frac{\nabla p}{\langle u_a\rangle}=\overline{R}_x \tag{11}$$

in which $\overline{R}_x$ is the effective flow resistivity for a single pore. In this way, the viscosity correction function for the fluid in a single pore with the statistically predicted size can be defined as

$$\overline{F}(\omega)=\frac{\langle\tau_a(\omega)\rangle}{\langle u_a(\omega)\rangle}\overline{R}_x^{-1}. \tag{12}$$

Thus Eq. (10) can be rewritten in the form

$$i\omega\overline{\rho}_x\langle u_a\rangle=\nabla p \tag{13}$$

with the effective dynamic density defined as

$$\overline{\rho}_x(\omega)=\rho_0\left(1-\frac{\overline{R}_x}{i\omega\rho_0}\overline{F}(\omega)\right). \tag{14}$$

In particular, expression (12) is consistent with the definition for the viscosity correction function

$$\overline{F}(\omega)=\frac{\langle\tau_a(\omega)\rangle}{\langle u_a(\omega)\rangle}\left[\frac{\langle\tau_a(0)\rangle}{\langle u_a(0)\rangle}\right]^{-1} \tag{15}$$

given by Yamamoto and Turgut[14] and by Attenborough[12] where it is treated as the normalized ratio of the average viscous force on the capillary walls per unit bulk volume to the average seepage velocity per unit bulk cross-sectional area.

On this basis, the complex compressibility of fluid in a single pore can be expressed as[5,13]

$$\overline{C}_x(\omega)=\frac{1}{\gamma P_0}\left(\gamma-\frac{\rho_0(\gamma-1)}{\overline{\rho}_x(N_{\rm pr}\omega)}\right), \tag{16}$$

where γ is the ratio of specific heats, P_0 is the atmospheric pressure, and $N_{\rm pr}$ is the Prandtl number.

Finally, the effects of the tortuosity and of the sample porosity Ω need to be included. We assume that $q\geqslant 1$ is the ratio of pore length to sample thickness so that q^2 is the tortuosity. In this case $\langle v_a\rangle=\Omega\langle u_a\rangle/q$ is the bulk velocity, the total volume flux divided by the sample cross-sectional area. Equation (13) holds with $\langle u_a\rangle$ replaced by $\langle v_a\rangle$ provided we replace $\overline{\rho}_x(\omega)$ by a bulk medium dynamic complex density $\rho_b(\omega)=(q^2/\Omega)\overline{\rho}_x(\omega)$. Thus

$$\rho_b(\omega)=\frac{q^2}{\Omega}\left(\rho_0-\frac{\Omega R_b}{i\omega q^2}\overline{F}(\omega)\right), \tag{17}$$

where $R_b=q^2\overline{R}_x/\Omega$ is the bulk flow resistivity. In this transformation the complex wave number for sound propagation in a single pore, $k_x(\omega)=\omega\sqrt{\overline{\rho}_x(\omega)\overline{C}_x(\omega)}$, is replaced by the bulk medium wave number

$$k_b(\omega)=qk_x(\omega). \tag{18}$$

TABLE I. Coefficients for the asymptotic expansions.

	Slit-like pores	Equilateral triangles	Circular pores
θ_1	$\frac{6}{5}e^{4\xi}-1$	$\frac{10}{7}e^{4\xi}-1$	$\frac{4}{3}e^{4\xi}-1$
θ_2	$\frac{1}{\sqrt{3}}e^{3/2\xi}$	$\sqrt{\frac{3}{5}}e^{3/2\xi}$	$\frac{1}{\sqrt{2}}e^{3/2\xi}$

Consequently, the complex impedance of the bulk medium is given by

$$Z_b(\omega)=\omega\rho_b(\omega)/k_b(\omega)=\sqrt{\rho_b(\omega)/C_b(\omega)}, \tag{19}$$

in which $C_b(\omega)=\Omega\overline{C}_x(\omega)$ is defined to be the bulk medium complex compressibility.

III. PADÉ APPROXIMATION

Experiments which relate to problems of geophysics[15,16] and outdoor sound propagation[17] suggest that in many granular materials the pore size can often be predicted closely by a log-normal statistical distribution (i.e., the logarithm of the size, $\log_2 s$, is normally distributed) with the probability density function

$$f(\phi)=\frac{1}{\sigma\sqrt{2}}e^{-[(\phi-\langle\phi\rangle)^2/2\sigma^2]}. \tag{20}$$

The statistical parameters $\langle\phi\rangle$ and σ can be selected in order to provide a tolerable fit to the experimentally determined cumulative pore size distribution $E(s)$, $E'(s)=e(s)$, and to cover a broad range of porous materials. It is assumed that for this log-normal distribution,

$$e(s)=f(\phi)\frac{d\phi}{ds},\quad s\geqslant 0, \tag{21}$$

where $\phi=-\log_2 s$ is the pore size in ϕ units, $\langle\phi\rangle=-\log_2\langle s\rangle$ is the mean pore size in ϕ units, and $\langle s\rangle$ is the median value of s. Of course, the same unit of length must be used for s as is used in the definitions of ϕ and $\langle\phi\rangle$, which is commonly expressed in mm.

It is obvious that even for the most simple pore geometries the adopted log-normal distribution will result in rather complex integral expressions which cannot be evaluated explicitly.[14] Although numerical integration techniques can be routinely used to calculate expressions (8) and (9), much more simple analytical approximations were recently proposed.[13] Relying upon the fact that the functional behavior of $\overline{F}(\omega)$ [see expression (12)] is described by the low- and high-frequency limit asymptotic developments it is possible to express the function in the following form:[13]

$$\overline{F}(\omega)\cong\frac{1+a_1\varepsilon+a_2\varepsilon^2}{1+b_1\varepsilon}, \tag{22}$$

where the coefficients in the rational approximation are

$$a_1=\theta_1/\theta_2,\quad a_2=\theta_1,\quad b_1=a_1. \tag{23}$$

The values of the coefficients θ_1 and θ_2 for three simple pore geometries are listed in Table I. The dimensionless quantity $\varepsilon=(-i\omega\rho_0q^2/R_b\Omega)^{1/2}$ is the equivalent of the dimension-

TABLE II. Values of the nonacoustic parameters used in the predictions.

Material	Flow resistivity, R_b, Pa s m^{-2}	Porosity, Ω	Tortuosity, q^2	Standard deviation, σ, ϕ units	Thickness, d,m
Glass beads	43 200	0.375	1.742	0.22	0.04
0.68 mm	(49 900)		(1.833)		
Glass beads	13 040	0.335	1.464	0.27	0.04
1.64 mm	(13 025)		(1.993)		
Coustone	31 520	0.40	1.664	0.31	0.0215
Foam YB10	1.52×10^5	0.61	1.918	0.44	0.0105

less parameter $\sqrt{-i}\lambda$ which is used by Attenborough.[12]

It is easy to show that the bulk flow resistivity is related to the median pore size as $R_b=(n\mu q^2/\langle s\rangle^2\Omega)\,e^{-2\xi}$, where $\xi=(\sigma\ln 2)^2$ and the shape parameter n depends on the assumed pore geometry ($n=3$ for slit-like pores, $n=8$ for circular pores, and $n=80$ for equilateral triangles). The value of the parameter n for realistic porous materials can be determined experimentally[18] and is unimportant for the process of acoustic modelling using the proposed formulation. The choice of the pore geometry is down to the ability to construct the analytical solution for $u_i(x,y,\omega)$ so that asymptotic expressions for the behavior of (12) in the low- and high-frequency limits can be determined. It has been shown that the effect of the pore geometry is relatively small in comparison with the effect of the pore size distribution.[12] Since the geometry of pores in realistic porous media are largely unknown, for practical applications of the above model it can be suggested to assume the circular pore geometry. For this geometry the statistical parameters of pore size distribution for many porous materials can be routinely obtained[19] and, as it will be shown further, used to model accurately their acoustic performance.

IV. MEASUREMENT OF NON-ACOUSTIC PROPERTIES OF POROUS GRANULAR MATERIALS

A representative selection of samples of consolidated and nonconsolidated granular mixes has been investigated. One of the consolidated samples has been manufactured from polyurethane foam granulates which are recovered from automotive waste. Another consolidated sample has been manufactured by CEP Acoustics Limited from flint particles consolidated with rubber-epoxy binder (*Coustone*). Nonconsolidated samples were graded mixes of spherical glass beads.

The following properties of porous samples have been determined experimentally: standard deviation of the pore size, σ, flow resistivity, R_b, porosity, Ω and tortuosity, q^2. The standard deviations have been measured using the water suction method.[19] This method also allows to determine the mean pore size and the internal pore surface area in consolidated and nonconsolidated porous samples.

The flow resistivity of loose and consolidated samples has been measured using the method detailed by Bies.[20] In the case of identical spherical glass beads of radius r_b the flow resisitivity can be predicted using the expression for the low-frequency limit of the drag parameter[21]

$$R_b=\frac{9\mu(1-\Omega)}{2r_b^2\Omega^2}\,\frac{5(1-\Theta)}{5-9\Theta^{1/3}+5\Theta-\Theta^2}, \tag{24}$$

where $\Theta\cong 0.675(1-\Omega)$.

The tortuosity has been derived from the upper-frequency data on the real part of the refraction index[22] and predicted in the case of identical spherical glass beads using the expression provided by Umnova[21]

$$q^2=1+(1-\Omega)(2\Omega)^{-1}. \tag{25}$$

The results of the measurements of the values of the nonacoustic parameters for the selected materials are compiled in Table II. The figures which appear in brackets correspond to the predicted values of the flow resistivity and tortuosity and are provided for comparison. Only the measured values were used for the prediction of the acoustic properties.

The results of the water suction experiments conducted on nonconsolidated mixes of glass beads support the assumption of the log-normal pore size distribution. Figure 1 shows the comparison between the experimental data for cumulative distribution of pore sizes in nonconsolidated mixes of glass beads and the theoretical fit provided by integrating expession (20) and minimizing the difference between the prediction and the experimental result for the cumulative distribution. The standard deviation in the size of pores in the stack of identical spherical glass beads is confined within the range between $0.22\leqslant\sigma\leqslant 0.27$ (ϕ units).

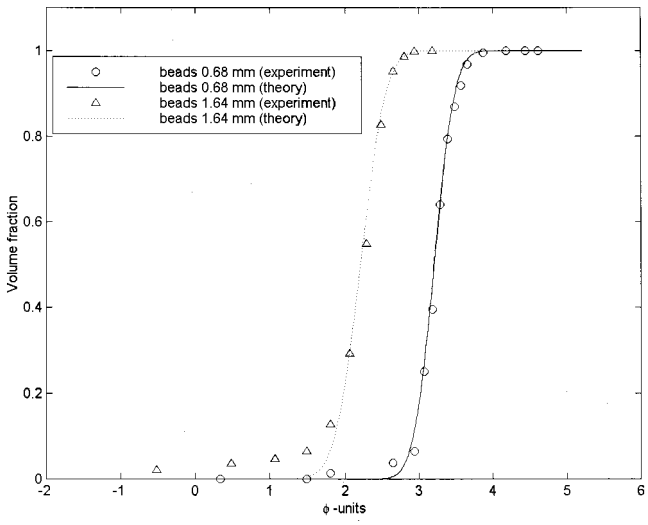


FIG. 1. Cumulative distribution of pore sizes in nonconsolidated samples of glass beads.

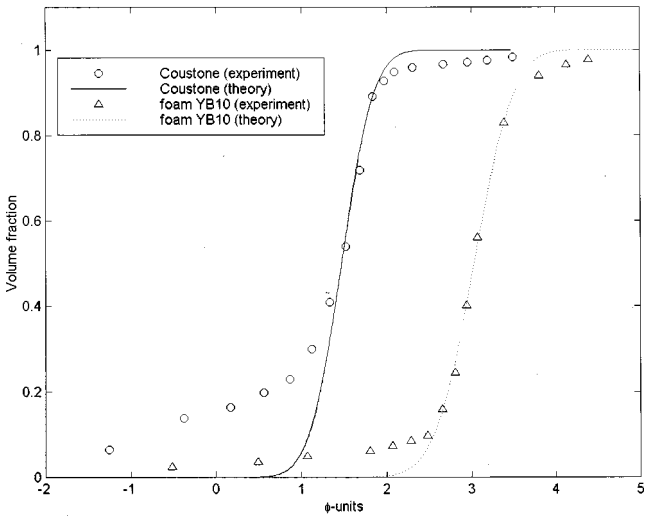


FIG. 2. Cumulative distribution of pore sizes in consolidated samples of Coustone and recycled PU foam.

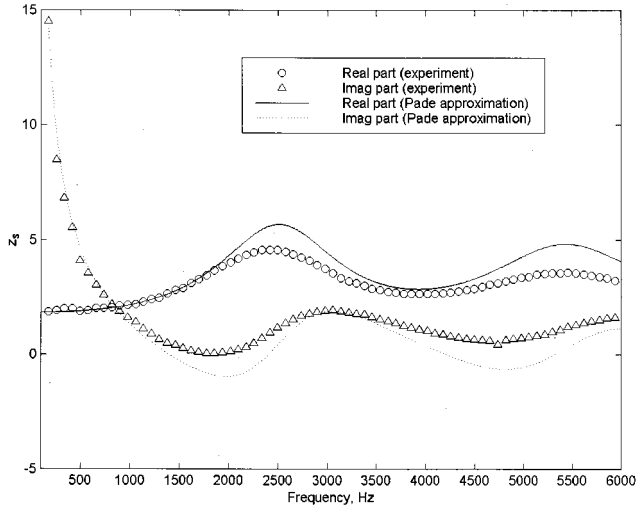


FIG. 3. The normalized surface impedance of a 40 mm layer of 0.68 mm nonconsolidated glass beads.

Figure 2 shows the comparison of the experimental data and the theoretical fit for the cumulative distribution of pore sizes in consolidated flint *Coustone* and in foam granulates YB10. The results suggest that the fit is close only above the predicted mean pore size for *Coustone* which is $\langle\phi\rangle = 1.713$. This disrepancy can be attributed to the consolidation effects and to the experimental errors. The consolidation process can considerably alter the geometry of the pores in loose granular mixes and the parameters of the pore size distribution.[18] The errors of a water suction experiment are associated with the dependence of the fluid contact angle on the material, the effect of gravity, and the effect of remaining water trapped in the closed pores at the end of the experiment. A comparison of results from recent water suction experiments conducted on graded glass beads with the theoretical predictions for the internal pore surface area suggest that the method can yield the accuracy of 6%–18%.[19] These results also show that the method is suitable for the measurement of the pore size distribution parameters of materials in which pore size ranges from 40 μm to 2.5 mm.

A better agreement between the experimental data and the theoretical fit for the cumulative pore size distribution is observed in the case of consolidated foam YB10 (Fig. 2). It will be shown, however, that the proposed acoustic model is not very sensitive to the deviation of the experimentally determined pore size distribution from the assumed log-normal law if the law provides the best possible fit to the experimental data.

V. MEASUREMENT OF THE ACOUSTIC PROPERTIES OF POROUS GRANULAR MATERIALS

The acoustic properties have been measured using a two-microphone impedance tube BK 4206 (ISO 10534-2) in the frequency range of 200–6000 Hz. With this setup the impedance of each sample was measured with and without an air gap, so that the characteristic impedance and refraction index could be determined. This setup is regularly calibrated. From our experience the reproducibility of the measurements of narrow-band absorption/surface impedance spectra is within 2%–5%. The impedance tube is installed in the vertical position to allow the acoustic properties of nonconsolidated mixes to be measured reliably. The normalized surface impedance, plane wave reflection coefficient and the absorption coefficient of a hard-backed porous layer are predicted using the following expressions:

$$z_s=(\rho_0 c)^{-1}Z_b\coth(-ik_b d),$$
$$R=\frac{z_s-1}{z_s+1}\quad\text{and}\quad\alpha=1-|R|^2, \tag{26}$$

respectively. The proposed Padé approximation (22) for circular pore geometry has been used to predict the complex wave number (18) and the characteristic impedance (19) so that a comparison between the experimental data and the predicted values of the surface impedance and the absorption coefficient could be made for the selected material samples. The results of this comparison are presented in Figs. 3–10.

Figures 3 and 4 show the results for a 40 mm hard-backed layer of 0.68 mm glass beads. Figure 3 shows the

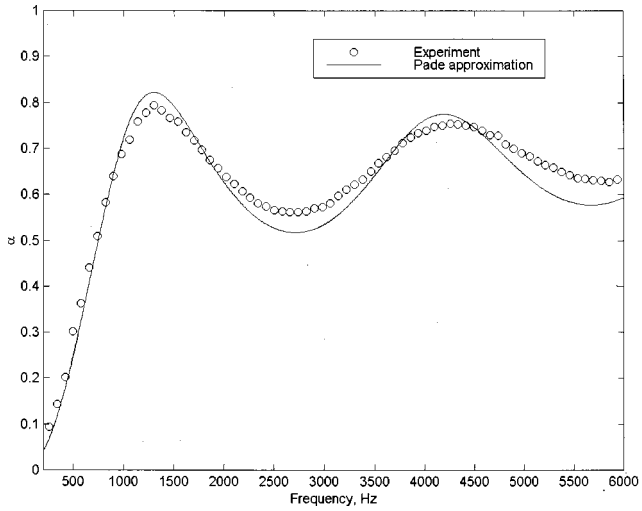


FIG. 4. The normal incidence plane wave absorption coefficient for a 40 mm layer of 0.68 mm nonconsolidated glass beads.

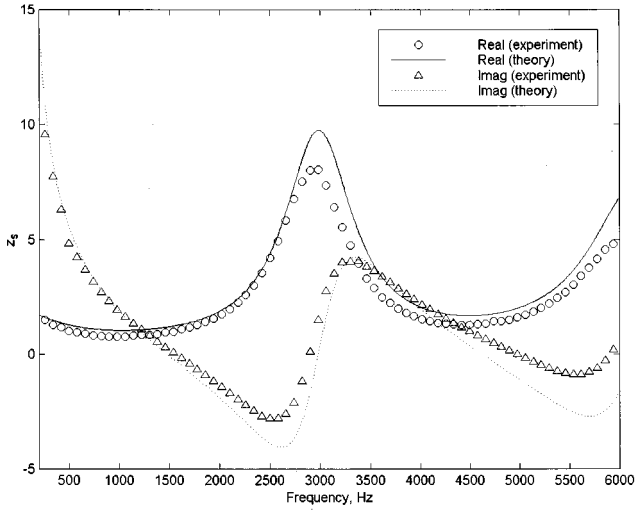


FIG. 5. The normalized surface impedance of a 40 mm layer of 1.64 mm nonconsolidated glass beads.

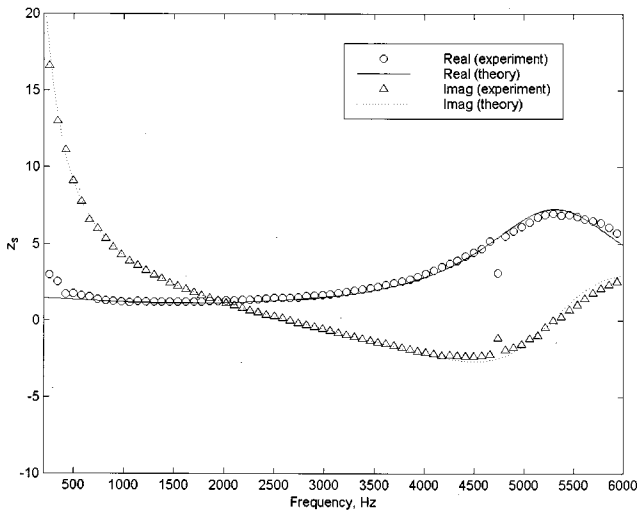


FIG. 7. The normalized surface impedance of a 20 mm layer of Coustone material.

values of the real and imaginary parts of the normalized surface impedance. The agreement between the experimentally determined values of the surface impedance and those predicted is reasonable throughout the considered frequency range. Some discrepancy is observed in the proximity of the acoustic resonances, which is likely to be associated with the deviation of the pore cross section in the material from the assumed cylindrical geometry. It is easy to demonstrate that variations from the assumed pore geometry can noticeably affect the imaginary part of the complex wave number and, therefore, the attenuation in the porous structure, resulting in the altered surface impedance. The measured and predicted absorption coefficient for this material is shown in Fig. 4. Good agreement is reported as a result of a close match between the measured and predicted surface impedance.

A better agreement between the measurement and the prediction is observed for the normalized surface impedance of a 40 mm hard-backed layer of 1.64 mm glass beads (see Fig. 5). In this case, the predicted results for the real and imaginary parts follow closely the experimental result practically throughout the considered frequency range. A limited discrepancy between the predicted imaginary part and the experimental result in the higher frequency range is likely to be caused by the deviation of the pore cross section in the material from the assumed cylindrical shape. A good agreement between the experimental result and the prediction is observed in the case of the absorption coefficient which is shown in Fig. 6.

Figures 7 and 8 show the normalized surface impedance and the plane wave absorption coefficient of a 21.5 mm thick hard-backed plate of *Coustone* material. In spite of a poor match between the experimentally determined pore size distribution and the assumed log-normal distribution there is excellent agreement between the measured and predicted results for the surface impedance (Fig. 7) and for the absorption coefficient (Fig. 8). *Coustone* is extensively used in architectural acoustics and noise control applications. Accurate modelling of the acoustic performance of this material is

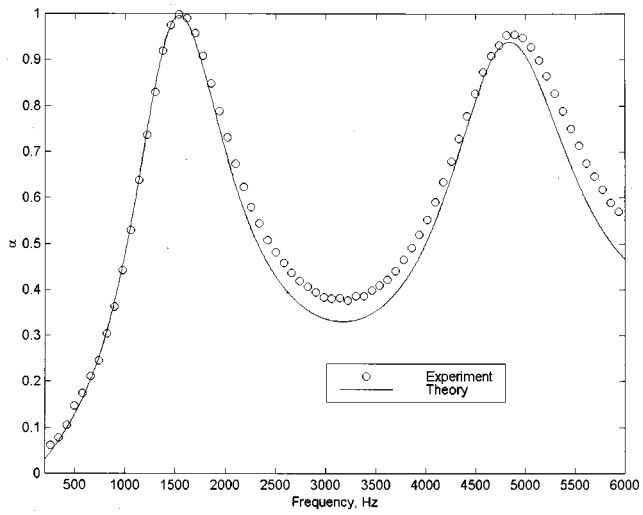


FIG. 6. The normal incidence plane wave absorption coefficient for a 40 mm layer of 1.64 mm nonconsolidated glass beads.

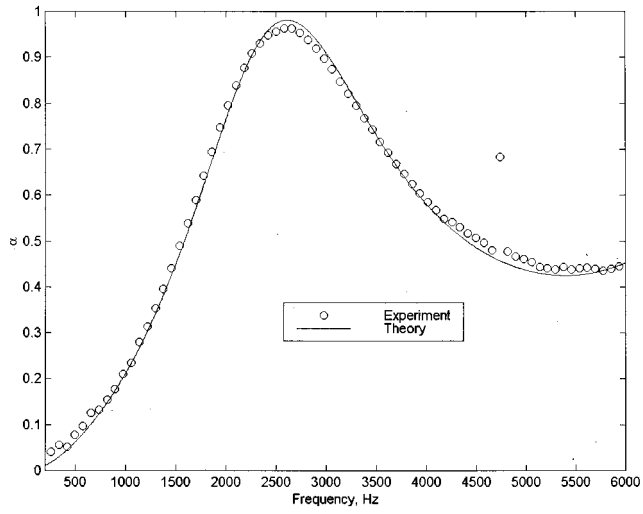


FIG. 8. The normal incidence plane wave absorption coefficient for a 20 mm layer of Coustone material.

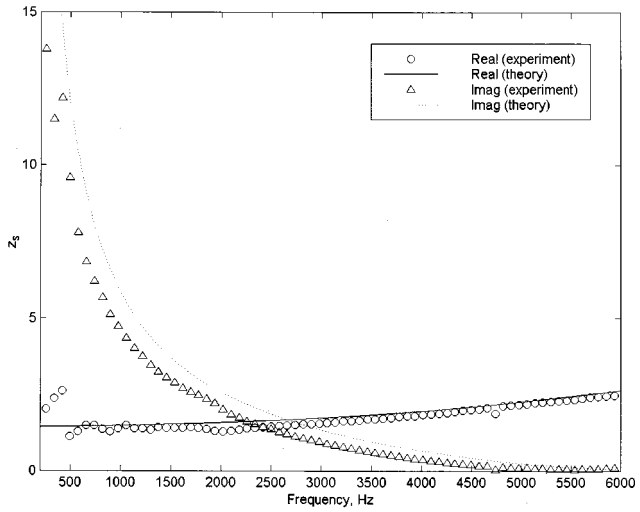


FIG. 9. The normalized surface impedance of a 10 mm layer of YB10 consolidated foam granulates.

important to optimize the required surface area, its thickness and the method of spatial distribution.

It has been discovered recently that the acoustic properties of consolidated materials manufactured from recycled granulated PU foams with complex surface geometry are similar or superior to those which are normally attributed to more common acoustic absorbers such as bulk foam and fiberglass. Analysis of the experimental data[18] and their comparison with theoretical results for the developed materials and for some common absorbers, suggest that it is the statistical properties of the pore size distribution which largely determine the effectiveness of the sound absorption of a porous structure. It is believed that unique statistical properties of the pore size distribution result from the composite porous structure of consolidated foam particles. Figure 9 shows the measured and predicted values of the normalized surface impedance of a 10.5 mm hard-backed layer of consolidated PU foam granulate YB10. There is an excellent agreement between the measured and predicted values for the real part of the surface impedance. The agreement for the imaginary part is less accurate in the low frequency range, which is likely to result from a considerable proportion of ''dead-end pores'' in this material. The size of these pores is not measurable by the water suction method, which is reflected in small discrepancy between the experimental and predicted values of the absorption coefficient shown in Fig. 10.

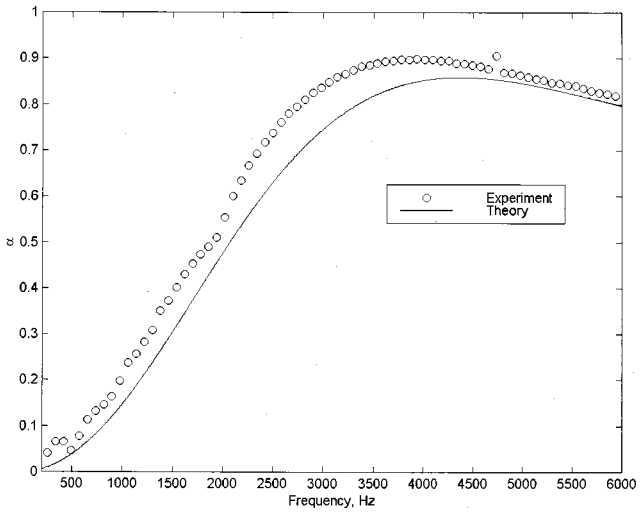


FIG. 10. The normal incidence plane wave absorption coefficient for a 10 mm layer of YB10 consolidated foam granulates.

VI. CONCLUSIONS

In this paper a computationally efficient rational (Padé) approximation for the acoustic properties of rigid frame porous media with a log-normal pore size distribution has been described. The model is based on the assumption that porous media can be modelled as a stack of parallel capillary tubes. This approach has been exploited in the past by many researchers and proved to provide good theoretical basis for the development of reliable models for the acoustic properties of acoustic porous materials. Unlike some other theoretical models which depend upon empirical shape factors, the proposed Padé approximation is based entirely on four measurable nonacoustic parameters. The water suction method used to determine the parameters of the log-normal distribution is nonacoustic, modelless and well adapted to acoustic materials and is, unlike the BET method, easy to reproduce in any basic acoustic laboratory requiring no expensive parts or chemicals.

The derived equations are accurate and robust for an extended range of values of microscopic material parameters, and can be used on materials with various pore geometries and a statistically distributed pore size. Careful theoretical consideration has been given in the model to the definition of the viscosity correction function describing viscous, thermal diffusion and pore size distribution effects on sound propagation inside a porous rigid frame material with three different pore geometries.

The model has been used to predict the acoustic properties of a representative selection of loose and consolidated porous granular materials. Close fit between the experimental data and the theoretical predictions has been observed for practical values of the nonacoustic parameters required by the model. The simple analytical form of the approximations means that these are of great practical interest for predictions of the acoustic field in the presence of a porous absorbing surface.

ACKNOWLEDGMENTS

The authors are grateful to the Engineering and Physical Sciences Research Council in the UK (Grant GR/L54905) for the support of this work. The authors would like to thank Dr. P. Leclaire for his help with the water suction experiments, the CEP Acoustics Limited for the provision of samples of *Coustone*, Professor D. C. Hothersall for very useful comments.

[1] C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, Amsterdam, 1949).
[2] S. Rayleigh, *Theory of Sound*, 2nd ed. (Dover, New York, 1976).
[3] M. A. Biot, ''Theory of elastic wave propagation in a fluid saturated porous solid,'' J. Acoust. Soc. Am. **28**, 168–191 (1956), Parts I and II.

[4] K. Attenborough, ''Acoustical impedance models for outdoor ground surfaces,'' J. Sound Vib. **99**, 521–544 (1985).
[5] M. R. Stinson, ''The propagation of plane sound waves in narrow and wide circular tubes, and generalisation to uniform tubes of arbitrary cross-sectional shape,'' J. Acoust. Soc. Am. **89**, 550–558 (1991).
[6] M. R. Stinson and Y. Champoux, ''Propagation of sound and the assignment of shape factors to model porous materials having simple pore shape geometries,'' J. Acoust. Soc. Am. **91**, 685–695 (1992).
[7] M. E. Delany and E. N. Bazley, ''Acoustical properties of fibrous absorbent materials,'' Appl. Acoust. **3**, 105–116 (1970).
[8] Y. Miki, ''Acoustical properties of porous materials-modifications of Delaney–Bazley models,'' J. Acoust. Soc. Jpn. **11**, 19–28 (1990).
[9] J.-F. Allard and Y. Champoux, ''New emperical equations for sound propagation in rigid frame fibrous materials,'' J. Acoust. Soc. Am. **91**, 3346–3353 (1992).
[10] D. K. Wilson, ''Relaxation-matched modeling of propagation through porous media, including fractal pore structure,'' J. Acoust. Soc. Am. **94**, 1136–1145 (1993).
[11] J. F. Allard, *Sound Propagation in Porous Media: Modeling Sound Absorbing Materials* (Elsevier, London, 1993).
[12] K. Attenborough, ''Models for the acoustical properties of air-saturated granular media,'' Acta Acust. (Beijing) **1**, 213–226 (1993).
[13] K. V. Horoshenkov, K. Attenborough, and S. N. Chandler-Wilde, ''Pade approximant for the acoustical properties of rigid frame porous media with pore size distribution,'' J. Acoust. Soc. Am. **104**, 1198–1209 (1998).
[14] T. Yamamoto and A. Turgut, ''Acoustic wave propagation through porous media with arbitrary pore size distributions,'' J. Acoust. Soc. Am. **83**, 1744–1751 (1988).
[15] E. L. Hamilton and R. T. Bachman, ''Sound velocity and related properties of marine sediments,'' J. Acoust. Soc. Am. **72**, 1891–1904 (1982).
[16] C. H. Juang and R. D. Holtz, ''Fabric, pore size distribution, and permeability of sandy soils,'' J. Geotech. Eng. Am. Soc. Civil Eng. **112**, 855–868 (1986).
[17] K. Attenborough, ''Outdoor sound propagation,'' Proceedings of the 16th International Congress on Acoustics, pp. 991–994, 1998, Seattle.
[18] M. J. Swift, Ph.D. thesis, University of Bradford, UK, 2000.
[19] P. Leclaire, M. J. Swift, and K. V. Horoshenkov, ''Determining the specific area of porous acoustic materials from water extraction data,'' J. Appl. Phys. **84**, 6886–6890 (1998).
[20] D. A. Bies and C. H. Hansen, ''Flow resistance information for acoustical design,'' Appl. Acoust. **13**, 357–391 (1980).
[21] O. Umnova, K. Attenborough, and K. M. Li, ''Cell model calculations of dynamic drag parameters in packings of spheres,'' J. Acoust. Soc. Am. **107**, 3113–3119 (2000).
[22] J. F. Allard, B. Castagnède, M. Henry, and W. Lauriks, ''Evaluation of tortuosity in acoustic porous materials saturated by air,'' Rev. Sci. Instrum. **65**, 754–755 (1994).

Efficient prediction methods for the micro-pressure wave from a high-speed train entering a tunnel using the Kirchhoff formulation

Taeseok Yoon
Department of Aerospace Engineering, Seoul National University, BD 136-318, Seoul National University, Seoul 151-742, Korea

Soogab Lee[a)]
IAAT (Institute of Advanced Aerospace Technology), School of Mechanical and Aerospace Engineering, Seoul National University, BD 301-1303, Seoul National University, Seoul 151-742, Korea

(Received 18 March 2001; accepted for publication 13 August 2001)

The compression wave generated by a high-speed train emerging from the exit portal of a tunnel gives rise to an impulsive noise called a micro-pressure wave. In this study, new methods for the prediction of sonic-boom noise are proposed. The first method combines acoustic monopole analysis and the method of characteristics with the Kirchhoff method. The compression wave from a train entering a tunnel is calculated by an approximate compact Green's function, and the resultant noise at the tunnel exit is predicted by a linear Kirchhoff formulation. The second method couples the Kirchhoff formulation with the Euler equation, which is solved numerically for the generation and propagation of the compression wave. Numerical prediction of the compression wave, the propagation in the tunnel, and the micro-pressure wave obtained by the present methods are compared with measured data. The numerical results exhibit a reasonable agreement with the experimental data. The proposed methods in this study are shown to be very useful design tools for the nose shape of trains and the geometry of tunnels, and they can be utilized to minimize the pressure fluctuation in the tunnel and the corresponding booming noise. [DOI: 10.1121/1.1409374]

PACS numbers: 43.50.Lj, 43.50.Pn, 43.28.Ra [MRS]

NOMENCLATURE

a_∞	speed of sound
a	height of tunnel
A	uniform cross-sectional area of tunnel
A_0	uniform cross-sectional area of train
A_T	variable cross-sectional area of train
b,D	width of tunnel ($=2b'$)
e	total energy per unit volume
$h_{\max}$	maximum train height
L	length of tunnel
l_p	length of profiled nose
l'	end-correction
M	train Mach number ($=U/a_\infty$)
$\mathbf{n}$	normal vector on the surface
p'	acoustic pressure
q	strength of volume source
Q	source strength
r	distance between a source position at the sound generation time and an observer at the sound receiving time
t	observer time
Δt	computational time step
U	speed of train
u,v,w	velocity components in x, y, z direction
x_p	measured or predicted position of pressure fluctuation in the tunnel
x, y, z	Cartesian coorinates
γ	ratio of specific heat
ϕ	velocity potential
ϕ^*	potential function
ρ_0	mean air density
ρ	density
θ	angle between the normal vector on the surface and the radiation vector
τ	retarded time ($=[t]$)
Ω	solid angle representing the configuration around the tunnel exit

I. INTRODUCTION

When a high-speed train enters a tunnel, a compression wave is generated ahead of the train, which propagates along the tunnel. The parameters affecting the magnitude and the gradient of the compression waves are train speed, blockage ratio, nose shape, etc. When the wave front of the compression wave arrives at the exit of the tunnel, a pressure pulse is radiated out, causing a strong impulsive noise, which becomes a serious environmental problem. This pressure pulse (or ''micro-pressure wave'') often results in structural damage of the tunnel and much annoyance for the population near the exit. The strength of the micro-pressure wave is

[a)]Electronic mail: solee@plaza.snu.ac.kr

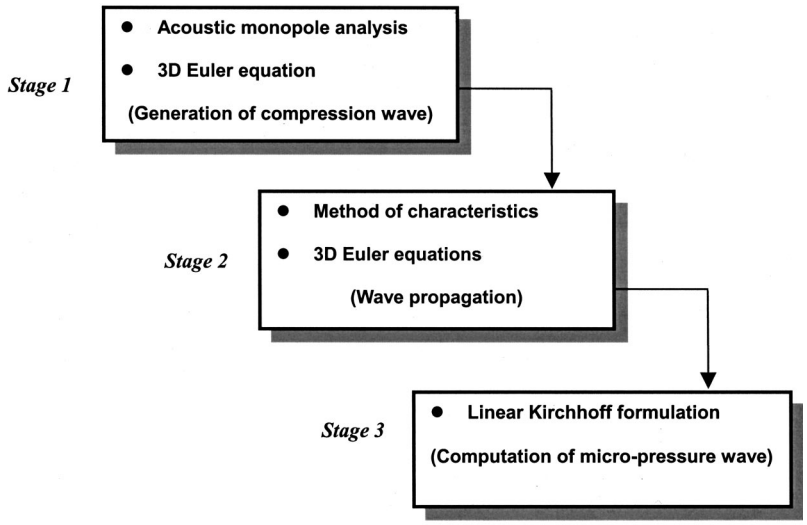


FIG. 1. Schematic of the prediction procedure for the micro-pressure wave (combined acoustic monopole analysis/method of characteristics–Kirchhoff method and CFD–Kirchhoff method).

proportional to the gradient of the incident compression wave. If a tunnel with a concrete slab track is longer than 3 km, a nonlinear steepening of the compression wave can produce a micro-pressure wave with peak magnitudes of over 50 Pa. The peak value of the micro-pressure wave depends on the train speed, the cross-sectional area ratio of train to tunnel, the train nose shape, and so on.[1]

The wave physics related to the micro-pressure wave can be divided into three phases:

(1) generation of the compression wave by a train entering a tunnel,
(2) propagation of the compression wave in tunnel, and
(3) emission of a micro-pressure wave at the tunnel exit.

In the current work, the formation of the compression wave is predicted by an empirical formula,[1] which uses parameters such as the train speed and the blockage ratio. The nose shape of the train, which plays an important role in formation of the compression wave, however, is not considered in the empirical formula. A sufficiently slender nose profile can reduce the micro-pressure wave strength by about 3 dB.[2,3] Therefore, the empirical formula from Ref. 1 in its original form is not adequate for the design procedure of train and tunnel. Howe analyzed the compression wave by using the exact acoustic Green's function for a semi-infinite and circular-cylindrical tunnel. He treated the nose of the train by a distribution of monopole sources along its symmetric axis, and calculated their interaction with the mouth of the tunnel.[4] He also derived the analytical representations of the Green's function describing the generation of sound waves within a semi-infinite flanged cylinder with sources located in the neighborhood of an open end using Rayleigh's method for the approximate calculation of out-going potential flow.[5,6] His approach was demonstrated for the flanged and cylindrical tunnels of semi-circular and rectangular cross sections, where exact analytic solutions of potential functions are not available. This analytical method can also consider arbitrary shapes of tunnels. As for the numerical approach to predict the compression wave, two- and three-dimensional Euler/Navier–Stokes solvers are used. Many approaches considering the effect of nose shape have attempted to analyze only the compression wave at the entrance of the tunnel.[7–9] Compared to other methods, this numerical approach requires much computational cost and time.

At stage (2), the propagation of the compression wave can be simulated by the method of characteristics using a one-dimensional Euler equation[10,11] or two/three-dimensional Euler/Navier–Stokes solvers. Only the case of a short tunnel (about 1 km) is adequate for two/three-dimensional Euler/Navier–Stokes calculations, considering the restriction of computer memory and time consumption.

For the prediction of the micro-pressure wave, the relationship between two waves (the micro-pressure wave and the compression wave) has been derived using the acoustic model of a vibrating circular piston on an infinite baffle plate.[1,12,13] Yoon *et al.* proposed a new method for the sonic boom noise using a linear Kirchhoff formulation which requires no approximation of frequency range. Furthermore, the method has an additional advantage in that it can predict the directivity of the micro-pressure wave.[14]

In this article, new methods to predict the micro-pressure wave are proposed. First, a CFD (computational fluid dynamics) technique is combined with a Kirchhoff formulation. An Euler finite difference solver is first executed to solve the near field. The resulting data are then transferred to the linear Kirchhoff formulation to predict the far field acoustics. The process is named the combined CFD–Kirchhoff method. The second method combines Howe's theory for the formation of the compression wave and the method of characteristics for the propagation of the compression wave with the Kirchhoff formulation. The method is named combined acoustic monopole analysis/method of characteristics–Kirchhoff formulation. Figure 1 summarizes the schematic of the above numerical and analytic calculation procedures.

The prediction results of the proposed methods are compared with each other and are validated with experimental data obtained in the Train Test Facility (T3F) at National Aerospace Laboratory (NLR) in the Netherlands.[14] Axisymmetric model trains and a tunnel of rectangular cross section

are used for the prediction and validation of the compression wave and the micro-pressure wave. The blockage ratio and other specifications will be discussed in Sec. V.

The present study is aimed at developing a useful design tool for the nose shape of high-speed trains and tunnel cross sections that can minimize the discomfort of passengers and mitigate the micro-pressure wave. The significance of the study reaches beyond just engineering interest in that the complex physical phenomena of aeroacoustics and aerodynamics are studied.

II. ACOUSTIC MONOPOLE ANALYSIS AND METHOD OF CHARACTERISTICS

A. Acoustic monopole analysis for formation of a compression wave

The velocity potential $\phi(\mathbf{x},t)$ of the sound generated by a distribution of volume source $q(\mathbf{x},t)$ is governed by the inhomogeneous wave equation

$$\frac{1}{a_\infty^2}\frac{\partial^2\phi(\mathbf{x},t)}{\partial t^2}-\frac{\partial^2\phi(\mathbf{x},t)}{\partial x^2}=-q(\mathbf{x},t). \tag{1}$$

The tunnel walls and surfaces of adjacent solid structures are assumed to be sufficiently rigid that the potential satisfies the wall boundary condition $\partial\phi/\partial x_n=0$ on all surfaces. The solution to Eq. (1) takes on the following form:

$$\phi(\mathbf{x},t)=-\int_V\int G(\mathbf{x},\mathbf{y};t-\tau)q(\mathbf{y},\tau)\,d^3\mathbf{y}d\tau. \tag{2}$$

Here, Green's function $G(\mathbf{x},\mathbf{y};t-\tau)$ is the solution of Eq. (1) if the right-hand side is replaced by the impulsive point source $\delta(\mathbf{x}-\mathbf{y})\delta(t-\tau)$. The integration in Eq. (2) is over the whole region of the fluid and over the time τ. Function G should satisfy both $\partial G/\partial x_n=0$ and $\partial G/\partial y_n=0$, where the field points $\mathbf{x},\mathbf{y}$ lie on the rigid tunnel boundaries and the solid surfaces outside the tunnel mouth.

When the tube is semi-infinite with an open end, the radiated sound wave resulting from the interaction of source q and the exit is determined by the following compact approximation of G:[5,6]

$$G(\mathbf{x},\mathbf{y};t-\tau)\approx\frac{a_\infty}{2A}\left\{H\left(t-\tau-\frac{|\phi^*(\mathbf{x})-\phi^*(\mathbf{y})|}{a_\infty}\right)-H\left(t-\tau+\frac{|\phi^*(\mathbf{x})+\phi^*(\mathbf{y})|}{a_\infty}\right)\right\}, \tag{3}$$

where $\phi^*(\mathbf{x})$ is the solution of Laplace's equation for incompressible and irrotational flow and $H(x)$ is the Heaviside unit step function. The function is normalized such that

$$\phi^*(\mathbf{x})\sim\begin{cases}x-\mathrm{l}' & \text{as } x\to-\infty \text{ in the tube,}\\ O(1/|\mathbf{x}|) & \text{as } x\to\infty \text{ outside the tube.}\end{cases}$$

In these equations, l' represents the "end-correction" of the tunnel mouth, which depends on the details of tunnel entrance geometry. The potential function $\phi^*(\mathbf{x})$ plays an important role in the calculation of sound generated by sources in the neighborhood of the open tunnel end provided the frequency is sufficiently small that the acoustic wave length is much larger than the tunnel diameter. Equation (3) is uniformly valid when regarded as a function of either $\mathbf{x}$ or $\mathbf{y}$, provided at least one of these points lies within the tunnel at a distance from the mouth large compared to the tunnel diameter.

The solution (3) determines the entry compression wave, prior to the propagation of wave in the tunnel, and is valid up to several tunnel diameters ahead of the train slightly after it enters the tunnel. The perturbation pressure p can be calculated from the linearized Bernoulli equation $p=-\rho_0\partial\phi/\partial t$. This yields the following form when Eq. (3) is substituted into Eq. (2):

$$p(\mathbf{x},t)=\frac{\rho_0 a_\infty}{2A}\int\{q(x'+U[t]-\mathrm{M}\phi^*(\mathbf{x}'),y',z')-q(x'+U[t]+\mathrm{M}\phi^*(\mathbf{x}'),y',z')\}\,d^3\mathbf{x}'. \tag{4}$$

Here, $[t]\approx t+(x-\mathrm{l}')/a_\infty$, and is the retarded time.

The source distribution q is nonzero only in the vicinity of the nose and tail where its cross-sectional area is changing. The compression wave is generated as the nose enters the tunnel and the length of the train should be assumed to be so large that the rear end can be ignored. If the train Mach number is so small that $O(M^2)$ and $M\phi^*(\mathbf{x}')$ terms are negligible during the formation of the compression wave, then

$$p\approx p(x,t)=\frac{\rho_0 U}{a}\int q(\mathbf{x}'+U[t],y',z')\frac{\partial\phi^*(\mathbf{x}')}{\partial x}d^3\mathbf{x}'. \tag{5}$$

Equation (5) describes the generation of a compression wave during the passage of the train nose into the tunnel. The monopole source distribution considering the nose shape of an axisymmetric train is as follows:

$$q(\mathbf{x}-\mathbf{U}t)=UA_0Q(x-Ut)\,\delta(y)\,\delta(z), \tag{6}$$

$$Q(x)=\frac{1}{A_0}\frac{\partial A_T(x)}{\partial x}. \tag{7}$$

Howe[5] suggested an extrapolation formula to extend the Mach number restriction. The prediction of Eq. (5) for very small Mach numbers can be modified to larger Mach numbers (about to 0.4) by the consideration of compressibility:

$$p(x,t)\approx\frac{\rho_0U^2A_0}{(1-\mathrm{M}^2)A}\int Q(x'+U[t])\frac{\partial\phi^*(x',0,z_T)}{\partial x'}d\mathbf{x}'. \tag{8}$$

In Eq. (5), the Rayleigh method for estimating the length of end-correction supplies an approximate representation of $\phi^*(x)$ for arbitrary shape of a given cross-sectional area of the tunnel. The function $\phi^*(x)$ is known analytically only for an unflanged and semi-infinite circular cylinder. Approximate representations of $\phi^*(x)$ for circular flanged ducts or rectangular cross-section can be obtained by the Rayleigh method.

For a rectangular tunnel with a flanged portal of height a and width b, the potential $\phi^*(x)$ describes potential flow from the rectangular duct of cross-section $2a\times b$ obtained from removal of the ground plane and introduction of the

image of the tunnel in $y=0$. Rayleigh's method is applied by taking $u(y,z)$, the exit plane velocity in the following form:[5]

$$u(0,y,z)=\frac{\partial \phi'}{\partial x}=\alpha\left(1+\frac{\mu_a y^4}{a^4}+\frac{\mu_b z^4}{b'^4}\right). \tag{9}$$

where $\alpha=1/[1+(\mu_a+\mu_b)/5]$.

The distribution of the potential function inside the tunnel $(x<0)$ is

$$\phi^*(x)=x-1'\sum_{n=1}^{\infty}\frac{8\alpha(-1)^n}{n^3\pi^3}\left(1-\frac{6}{n^2\pi^2}\right)\times\left\{a\mu_a\cos\left(\frac{n\pi y}{a}\right)e^{n\pi x/a}+b'\mu_b\cos\left(\frac{n\pi z}{b'}\right)e^{n\pi x/b'}\right\}. \tag{10}$$

Function $\phi^*(x)$ outside the tunnel $(x>0)$ is given as follows:

$$\phi^*(x)=\frac{-1}{2\pi}\int_{S_0}\frac{u(y',z')dy'dz'}{\sqrt{x^2+(y-y')^2+(z-z')^2}}. \tag{11}$$

The coefficients of μ_a and μ_b are determined to minimize the length of end-correction by Kelvin's minimum energy principle. The value of μ_a and μ_b will be mentioned later in Sec. V.

B. Method of characteristics for propagation of compression wave

The micro-pressure wave depends on the waveform of the incident compression wave at the tunnel exit. The magnitude of the micro-pressure wave is approximately proportional to the time derivative of the pressure of the incident compression wave $(\partial p/\partial t)$. The micro-pressure wave becomes larger in a long tunnel with a concrete slab track because the wave front of the compression wave becomes steep due to a nonlinear effect during the propagation. In a long ballasted track tunnel, the wave front of the compression wave does not steepen and the pressure gradient of the wave front decreases during propagation. The maximum value of the micro-pressure wave in long ballasted track tunnels becomes smaller than that in short tunnels.[1]

The effect of a ballasted track is attributed to its porosity. A ballasted track can be regarded as a porous sound-absorbing material, which is effective in the low-frequency range. Penetration of air into the ballasted track causes a dissipation effect and prevents the wave front of the compression wave from steepening.

The hyperbolic wave equation for the forward-moving compression wave in a one-dimensional sense is as follows:

$$\frac{\partial u}{\partial t}+\left\{(1-\alpha_c)a_\infty+\frac{\gamma+1}{2}u\right\}\frac{\partial u}{\partial x}=\frac{1}{2}\left\{f_1+\nu^*\frac{\partial^2 u}{\partial x^2}\right\}, \tag{12}$$

where u is the medium air velocity accompanying the compression wave, f_1 is the frictional force at the periphery of the tunnel, γ is the specific heat ratio of air, and x is the spatial coordinate along the tunnel. The coefficients α_c $(=1.4\times10^{-3})$ and $\nu^*(=4.35\times10^{-5})$ of Eq. (12) represent the effect on propagation velocity of compression wave and the dissipation effect, respectively.[15]

III. CFD FOR FORMATION AND PROPAGATION OF COMPRESSION WAVE

A. CFD technique

An unsteady, compressible, and three-dimensional Euler equation is solved to analyze the flow field around the high-speed train. The system of equations consists of a local time derivative term and three convective flux vectors. In physical coordinates, the governing equation is

$$\frac{\partial Q}{\partial t}+\frac{\partial E}{\partial x}+\frac{\partial F}{\partial y}+\frac{\partial G}{\partial z}=0, \tag{13}$$

where

$$Q=\begin{pmatrix}\rho\\ \rho u\\ \rho v\\ \rho w\\ \rho e\end{pmatrix},\quad E=\begin{pmatrix}\rho u\\ \rho u^2+p\\ \rho uv\\ \rho uw\\ (\rho e+p)u\end{pmatrix},$$

$$F=\begin{pmatrix}\rho v\\ \rho uv\\ \rho v^2+p\\ \rho vw\\ (\rho e+p)v\end{pmatrix},\quad G=\begin{pmatrix}\rho w\\ \rho uw\\ \rho vw\\ \rho w^2+p\\ (\rho e+p)w\end{pmatrix}.$$

The equation of state is

$$e=\frac{1}{\rho(\gamma-1)}p+\frac{1}{2}(u^2+v^2+w^2). \tag{14}$$

Roe's finite volume flux difference splitting technique, based upon the solution of Riemann problem[16] is used for spatial discretization and MUSCL (monotone upstream-centered scheme conservation laws) with a van Albada flux limiter used to achieve third-order spatial accuracy. Calculations of unsteady flow field around a moving body require a time accurate numerical integration. In the present study, Yoon's LU-SGS scheme[17] is chosen for efficient time marching due to the huge demand on computer power.

B. Grid systems

The grid system for computation of the three-dimensional train/tunnel aerodynamic interaction is characterized by a moving body confined to linear motions on the rail, relative motions between the solid body and the background grid, ground proximity, and large computational domains through which a train moves.

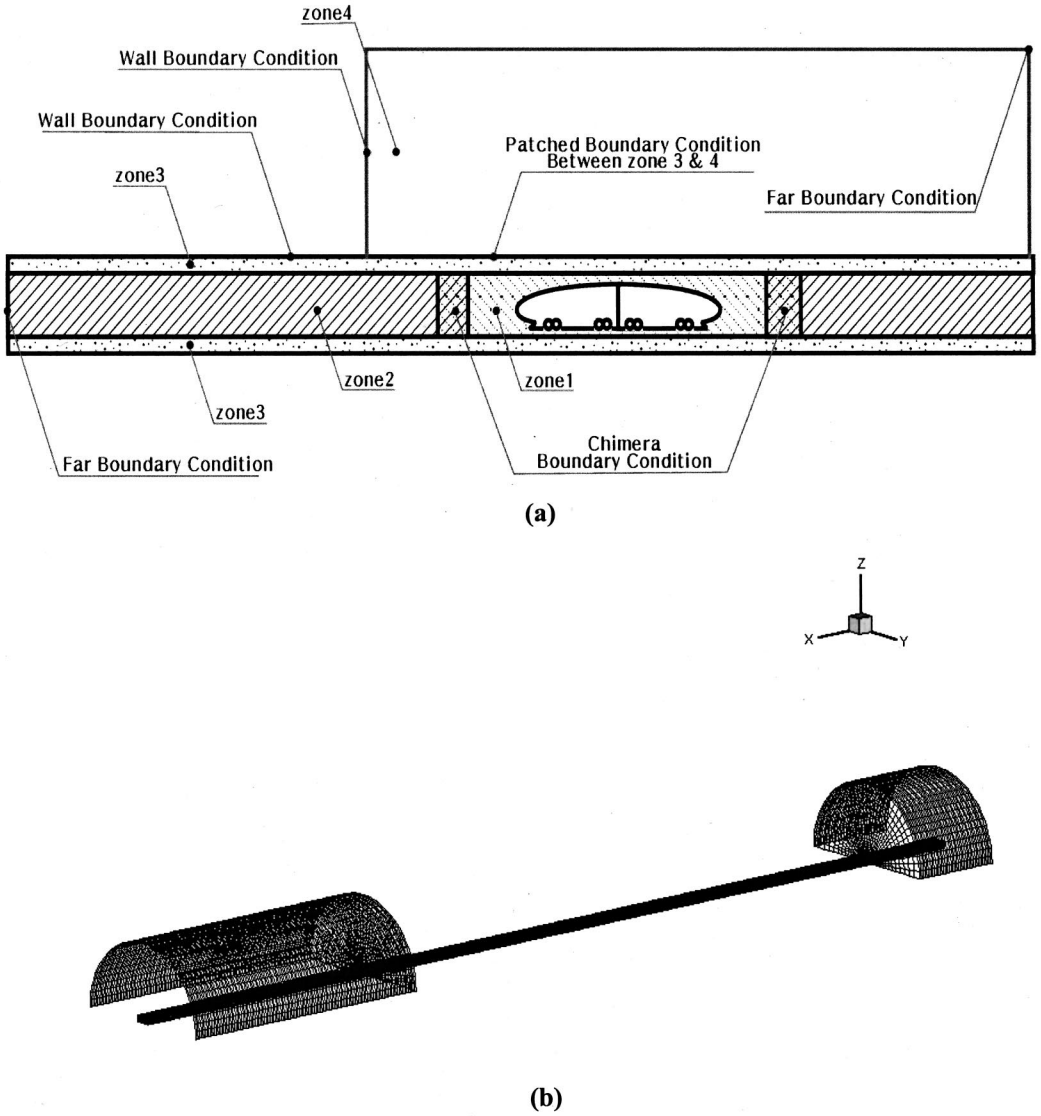


FIG. 2. (a) Schematic diagram for zone decomposition and boundary condition and (b) five-block grid system for the CFD calculation.

To accommodate the relative motion, Fujii[9,18] used FSA (fortified solution algorithm) while Mestreau[19] used an unstructured grid with automatic remeshing. Fujii's approach for the current problem is similar to the Chimera[20] procedure composed of hole construction and linear interpolation at the boundaries. To obtain a stable solution, a moderate cell volume ratio between a giving cell and a target cell has to be maintained at the boundaries. If the region swept by the train is large, a huge number of grid points along the region is required in order to preserve the proper interpolation criterion. Fujii used an intermediate zone to overcome this requirement, but this method made the linear interpolation routine more complicated. In this study, three types of DDT (domain decomposition technique) are applied. They are multi-block, patched, and overlapping grids to obtain both computational efficiency and extensibility.

TABLE I. Number of grid points used for prediction flow field.

Zone	Grid (x,y,z)	Sum of grid
1	(71,23,25)	40 825
2	(431,23,25)	247 825
3	(431,23,5)	49 565
4	(120,18,25)	54 000
5	(40,18,25)	18 000
Total sum of grid		410 215

Figure 2 shows the schematic zone decomposition and the zonal interface for this problem. Zone 1 moves with the high-speed train while zone 2 is fixed. Consequently, a simplified Chimera hole construction and linear interpolation are applied at the fore and aft overlapping regions of the moving train. As no stiff gradient of the flow variables exists except in the vicinity of the train, and as the wave phenomena in the tunnel are almost one-dimensional, linear interpolation is sufficient without adopting a tedious and complicated three-dimensional conservative treatment.[21] Zone 3 is an intermediate zone sharing a sliding surface with zone 1. Zone 4 is the tunnel entrance zone. At the tunnel exit, zone 5 similar to zone 4 can be defined. Table I lists the number of grid points used for prediction of near field flow. In the computations, a constant computational time step of $\Delta t = 0.01$/iteration is used. It takes approximately 16 s per iteration at 450 Mhz

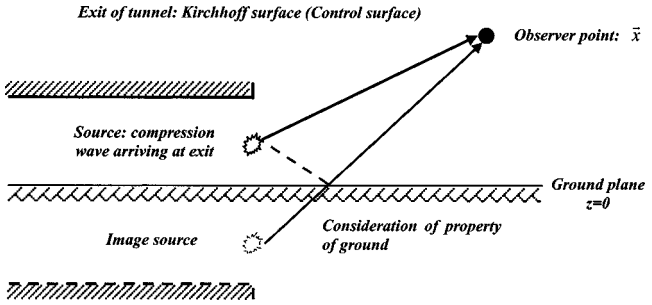


FIG. 3. Kirchhoff (control) surface for prediction of the micro-pressure wave and consideration of the effect of ground (image source of sound).

with a digital Alpha CPU. The length and time scales are normalized by D and D/a_∞, respectively, where D is the width of the tunnel.

IV. KIRCHHOFF FORMULATION FOR RADIATION OF MICRO-PRESSURE WAVE

A linear Kirchhoff formulation is applied to predict the micro-pressure wave at the tunnel exit. The near field flow data obtained by acoustic monopole analysis/method of characteristics or CFD are used as an input to Kirchhoff formulation. The Kirchhoff integral formula for a stationary control surface is the analytic expression of Huygens' principle.[22] The Kirchhoff formulation is as follows:

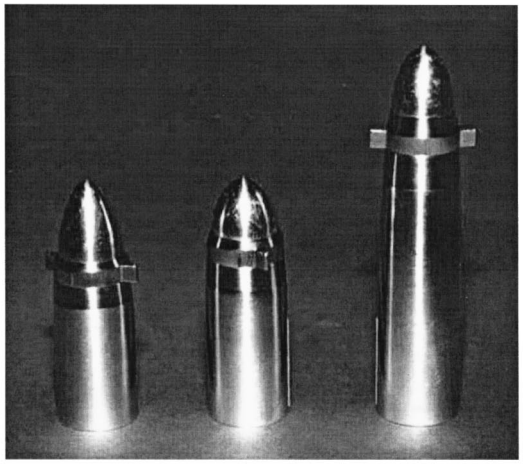
$$p'(\mathbf{x},t)=\frac{1}{4\pi}\int_S\int\left[\frac{\cos\theta}{r^2}p'-\frac{1}{r}\frac{\partial p'}{\partial n}+\frac{\cos\theta}{a_\infty r}\frac{\partial p'}{\partial \tau}\right]dS(\mathbf{y},\tau). \quad (15)$$

Here, p' is the perturbed pressure, ($\mathbf{x}$, t) are the observer's location and time, ($\mathbf{y}$, τ) are the source location and retarded time variables, θ is the angle between the normal vector ($\mathbf{n}$) on the surface and the radiation vector ($\mathbf{r}$), and r is the distance between an observer and a source at the retarded time. Note that pressure and its derivatives are calculated at the retarded time, τ. Equation (15) has been widely used for aeroacoustic predictions of high-speed impulsive noise and transonic blade-vortex interaction noise generated by a rotor.[23–25]

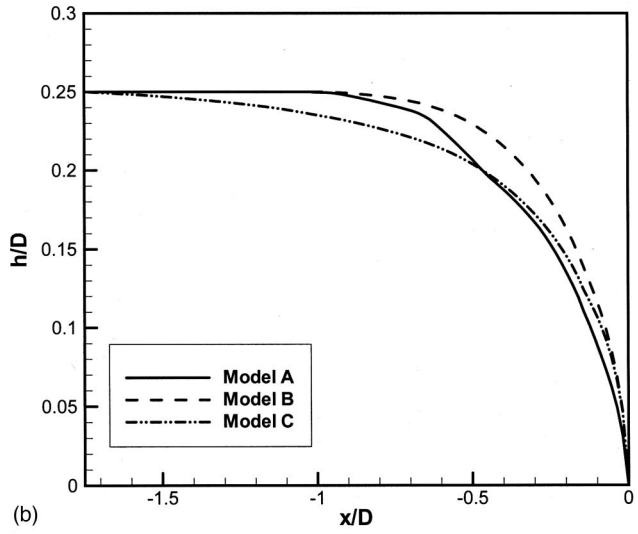
The integration is performed at the control surface, called the Kirchhoff surface, containing the flow field information. In the present study, the exit plane normal to the tunnel exit is considered as the control surface, and the image source of sound is imposed to consider the ground effect. The coefficient of sound absorption varies with the frequency of the incident wave and the material property.[26] Us-

TABLE II. Test types used in prediction and validation.

Test type	Model train (mm^2)	Tunnel area (mm^2)	Blockage ratio (%)	Aspect ratio (h_{max}/l_p)
A (K-TGV)	314.2	3680	8.5	10 mm/40 mm
B (Short nose)	314.2	3880	8.1	10 mm/40 mm
C (Long nose)	314.2	3880	8.1	10 mm/70 mm



(a)



(b)

FIG. 4. (a) Axisymmetric model trains A, B, and C and (b) a sketch of the model train.

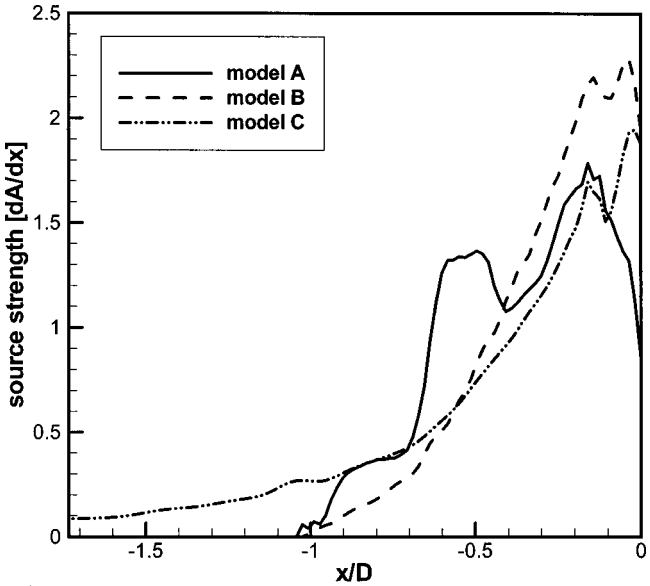


FIG. 5. Source strength of model trains A, B, and C.

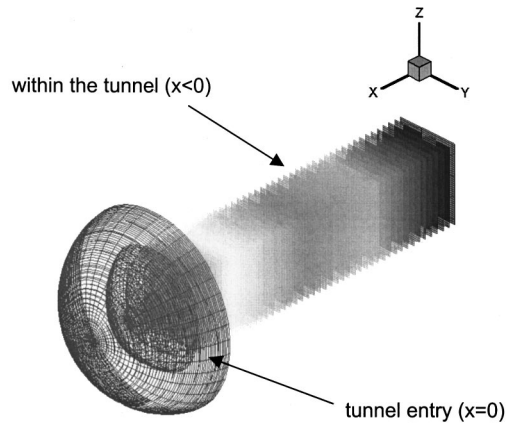


FIG. 6. Distribution of the potential function $\phi^*(\mathbf{x})$ within the tunnel and outside the tunnel.

ing the coefficient, the absorption of the acoustic wave by the ground is calculated. In the present study, the rise time of the compression wave is regarded as the characteristic time, and the absorption coefficient is determined with respect to this time. Figure 3 shows the schematics of the image source and the Kirchhoff surface under consideration. At the tunnel exit portal, the baffle plate is not installed.

V. RESULTS

A. Formation of the compression wave

All model trains have a dimension of 20 mm in diameter and a length of 2340 mm. They are used to verify two different prediction methods. One is acoustic monopole analysis and the other is CFD.

The tunnel has a rectangular shape of constant height (40 mm) and width (97 mm), corresponding to a blockage ratio of 8.1% except model A.[14] Cross-sectional areas of each tunnel and train model with its blockage ratio are listed in Table II. The track structure has the property of a ballasted track.

The blockage ratio used in prediction and experiment is very small compared to those in Japan or France. The prediction model is based upon the standards of K-HST ("Korea-High Speed Train") that will go into service in 2004. The tunnel length is 7640 mm. Figures 4(a) and (b) present the nose geometry and sketches of the three models: trains A, B, and C. Model train C is the most slender in shape. Corresponding strengths of sources (dA/dx) are plotted in Fig. 5.

The potential function distribution by Eqs. (9)–(11) is shown in Fig. 6. In the region $x>0$, namely outside the tunnel, the flow spreads hemispherically. The function $\phi^*(x)$ varies continuously through the tunnel entrance, increasing from negative values at $x<0$. In this calculation, as mentioned in Sec. II, the value of μ_a and μ_b are determined to minimize the length of end-correction by Kelvin's minimum energy principle.[5] For height (a) 40 mm and width (b) 97 mm, μ_a is 0.911 and μ_b is 0.820, respectively, and the minimized end-correction length (l') is 1.008.

At a speed of 350 km/hr (M=0.28), the entry compression wave and its pressure gradients for the three models A, B, and C are predicted by the analytic theory of Eq. (8), the result of which is plotted against nondimensional time $U[t]/a$ in Fig. 7. As shown in Table II, the blockage ratio of the model train A is slightly larger, about 0.4%, than that of model trains B and C. However, in the prediction of the compression wave, the same blockage ratio was used to investigate the effect of the train nose shape. As all the models have the same blockage ratio (8.1%), the maximum pressure rise is also the same. As for the gradient of the compression wave, however, model C has the minimum gradient. Although models A and B have the same profile nose length (l_p), the gradient of the compression wave of model B, which is relatively more blunt, is greater than that of model A. Since the magnitude of the micro-pressure wave is proportional to the pressure gradient of the compression wave arriving at the tunnel exit,[1] model B is to exhibit the micro-pressure wave of the largest amplitude.

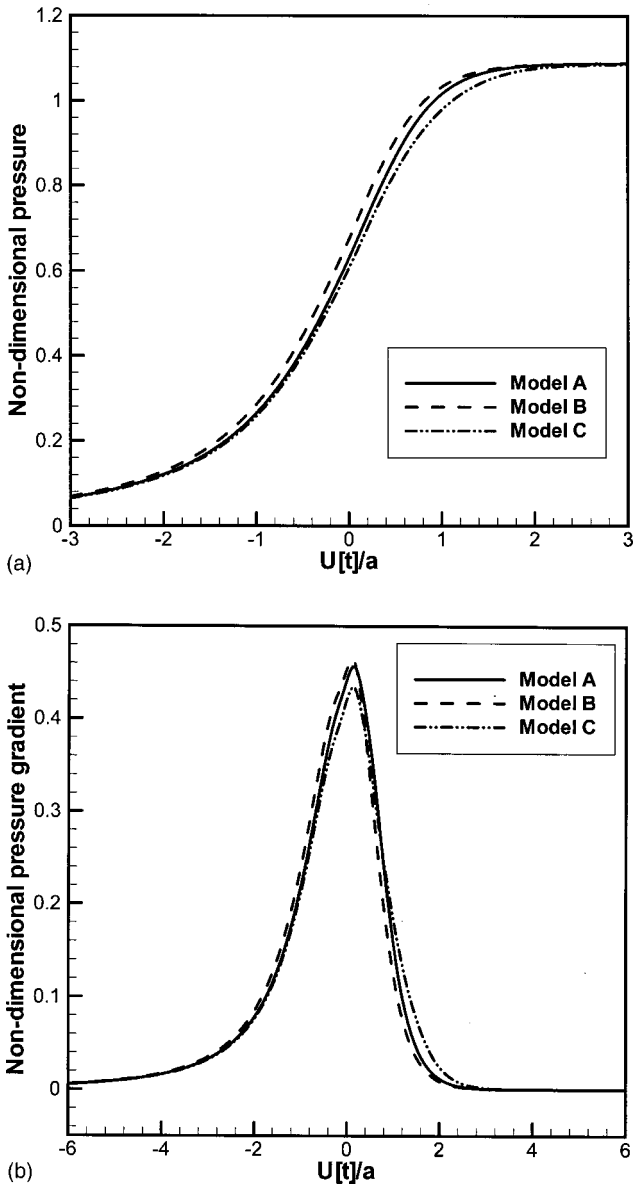
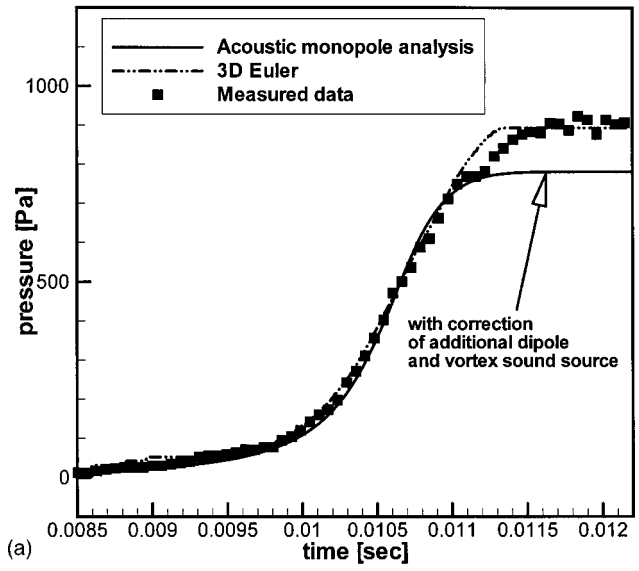


FIG. 7. (a) Entry compression wave and (b) pressure gradient for model trains A, B, and C at Mach number of 0.28 for the same blockage ratio of 8.1%. (Pressure and pressure gradient are normalized by $\rho_0 U^2 A_0/A$ and $\rho_0 U^3 A_0/aA$, respectively.)



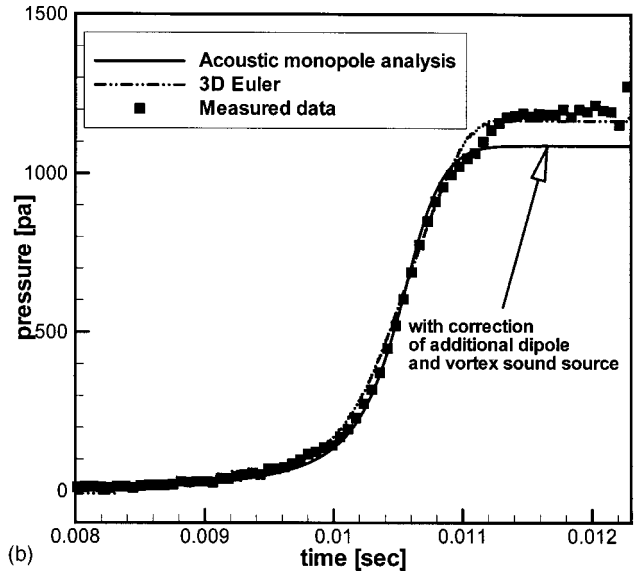


FIG. 8. Comparison with measured and predicted entry compression waves obtained by acoustic monopole analysis with second-order dipole correction and an Euler calculation for (a) 300 km/hr and (b) 350 km/hr.

To validate the prediction method, the entry compression waves at position 4.3% ($x_p/L=4.3\%$) of whole tunnel length (7640 mm) from the tunnel entry obtained by analytical theory and direct Euler computation are compared with the experimental data in Figs. 8(a) and (b). Model C is presented as a representative case. The train speed is (a) 300 km/hr and (b) 350 km/hr. Corrections to the monopole theory can be invoked to render theoretical predictions compatible with experiment.[27] First, an additional dipole source representing the compression wave drag acting on the nose of the train can be considered. The effect of this is to increase the amplitude predicted by the monopole source by a factor of $(1+A/A_0)$, where A/A_0 is the cross-sectional ratio of train to tunnel. Second, a dipole source is attributed to vortex sound sources in the shear layer of the back-flow out of the tunnel displaced by the train. This gives a progressive increase of about 1%. The measured[14] and predicted results show good agreement with each other. The prediction result

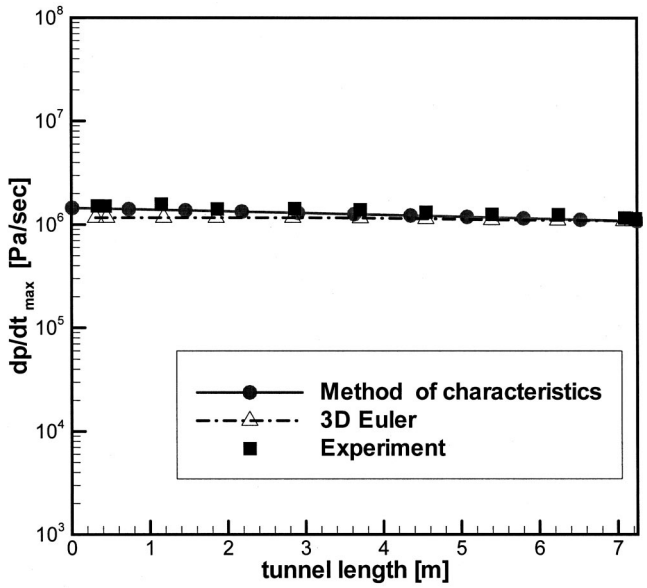


FIG. 9. Comparison with the maximum pressure gradient in the tunnel for model train C at a speed of 350 km/hr.

of the acoustic monopole with the correction of additional dipole and vortex sound source differs slightly in the maximum amplitude of pressure wave with the measured data but the gradient of the compression wave, playing an important role in generating the micro-pressure wave, approximately corresponds to the measured data. An accurate prediction of the compression wave is very important in the prediction of the micro-pressure wave because its strength is proportional to the gradient of the compression wave. In the viewpoint of computing time, the acoustic monopole method can markedly reduce the time consumption compared with the Euler simulation.

B. Propagation of the compression wave

The entry compression wave obtained by acoustic monopole analysis is used as an input to the method of characteristics. The magnitude of maximum pressure gradient in the tunnel by the method of characteristics is compared to experimental data in Fig. 9 in the case of train speed 350 km/hr for model C. The model tunnel having a length of 7640 mm is equivalent to a 1.2-km short tunnel in actual scale. Therefore, the influence of ballasted track on dissipation and dispersion of propagating compression wave is not so evident. Thus, for the prediction of propagation using the Euler equation, the track effect is not considered. It is also troublesome to consider the damping effect in the Euler equation. For a more accurate calculation, it is necessary to use a Navier–stokes solver with porous modeling on the track. As mentioned previously in Sec. I, it is appropriate to use the Euler equation in the prediction of train/tunnel aerodynamic interaction for a short tunnel because the damping effect can be regarded as negligible.

The propagation of the compression wave along the tunnel is simulated by three-dimensional Euler equation as shown in Figs. 10(a)–(d) for model train C with an entrance speed of 350 km/hr. In Fig. 10(a), it is observed that the

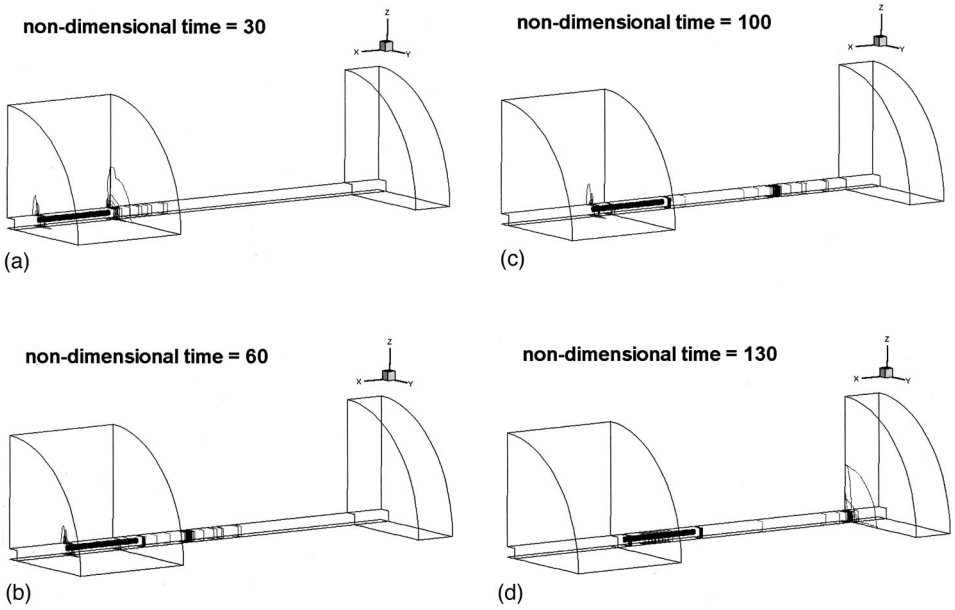


FIG. 10. Time series of the pressure contour representing the propagation of the compression wave in the case of model C having a speed of 350 km/hr.

compression wave is generated from the head of the train entering the tunnel. The compression wave propagates towards the exit of the tunnel faster than the train itself at about the speed of sound: Figs. 10(b) and (c). Finally, in Fig. 10(d), the compression wave radiates from the exit as a pulselike wave called the micro-pressure wave. Additionally the expansion wave is created at the entrance when the rear end of the train arrives at the entry of the tunnel. The predicted pressure fluctuation in the tunnel obtained by the three-dimensional Euler equation is compared with the measured data in the vicinity of the tunnel exit in Fig. 11. The measurement position is at $x_p/L=94.6\%$ from the tunnel entry. The predicted compression wave shows good agreement with the experimental data in both waveform and pressure gradient. The maximum pressure gradients in the tunnel obtained by the three-dimensional Euler equation are shown in Fig. 9. The result shows that the maximum value of pressure

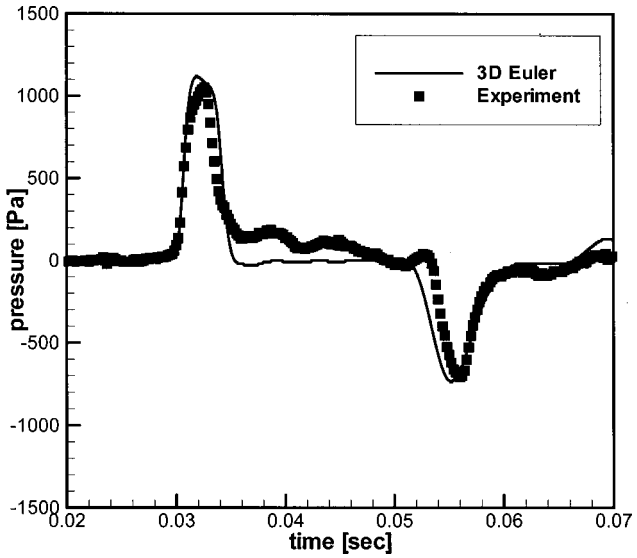


FIG. 11. Comparison of measured and predicted pressure fluctuations in the vicinity of the tunnel exit ($x_p/L=94.1\%$) obtained by 3D Euler equation for model train C at speed of 350 km/hr.

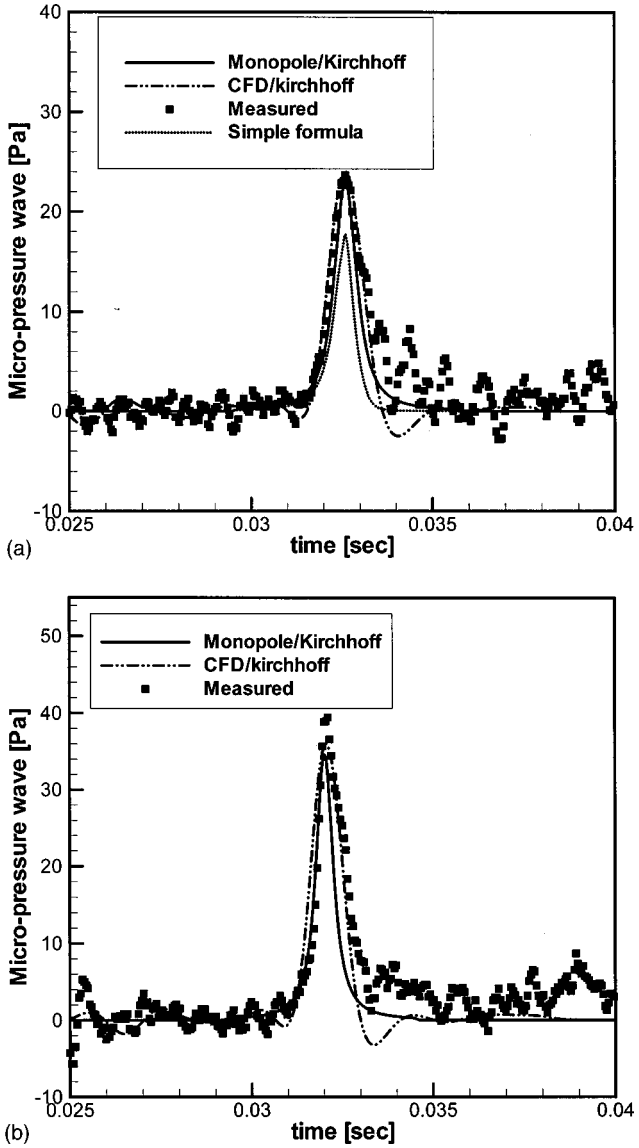


FIG. 12. Comparison of predicted and measured micro-pressure waves for (a) 300 km/hr and (b) 350 km/hr in the case of model C.

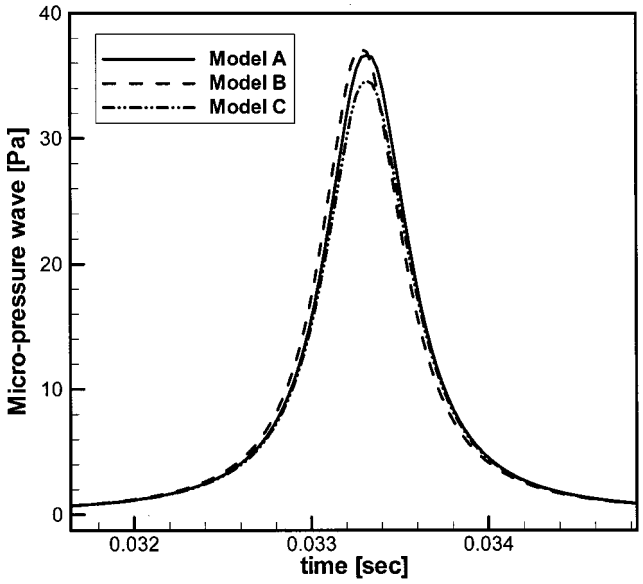


FIG. 13. Effect of the nose shape of the train on the micro-pressure wave at a speed of 350 km/hr and blockage ratio 0.081, obtained by prediction using combined acoustic monopole analysis/method of characteristics–Kirchhoff formulation.

gradient at exit (dp/dt_{EXIT}) is nearly equal to that of entry (dp/dt_{ENTRY}). Even though the experimental result shows dissipation during propagation, there is no large dissipation as can be observed in a long tunnel with ballasted track. The damping effect can thus be regarded as negligible due to the property of the short model tunnel in the Euler calculation.

C. Generation of the micro-pressure wave

The flow-field data obtained by the acoustic monopole analysis/method of characteristics and CFD is transferred to a linear Kirchhoff formulation to predict the micro-pressure wave at the exit of tunnel. Figure 12 shows the comparison of calculated results with experimental data at speeds of 300 and 350 km/hr for model train C. The micro-pressure wave is measured and predicted at 120 mm from the exit and 20 mm above the ground, inclined 45 degrees to the tunnel. For a combined acoustic monopole analysis/method of characteristics–Kirchhoff formulation, a second-order dipole correction to the monopole theory is applied as shown in Figs. 8(a) and (b). The predicted micro-pressure wave shows satisfactory agreement in both amplitude and waveform. Therefore, the proposed prediction method, a combined acoustic monopole analysis/method of characteristics–Kirchhoff method and CFD–Kirchhoff method, works very well in predicting the micro-pressure wave as well as the aerodynamic interaction between the train and the tunnel.

In previous research,[1] the following simple relationship between the micro-pressure wave and the compression wave arriving at the exit portal using far field and low-frequency approximations were used:

$$p'(\mathbf{x},t) \sim \frac{A}{\Omega a_\infty r}\left[\frac{\partial p}{\partial t}\right]_{\mathrm{EXIT},t-r/c}, \tag{16}$$

where Ω is the effective solid angle into which the wave radiates; r is the distance from the exit portal to the measured point. The dotted line in Fig. 12(a) represents the prediction result from this simple formula. In prediction by the simple formula, the pressure gradient at the tunnel exit is obtained by the acoustic monopole analysis and method of characteristics. The image source is also used to consider the ground effect as in the Kirchhoff formulation. Equation (16) underpredicts the magnitude of the micro-pressure wave by about 30%.

To obtain the entry compression waveform and the resultant micro-pressure wave, it takes 67.2×10^3 s for a combined CFD–Kirchhoff calculation, whereas a combined acoustic monopole analysis/method of characteristics–Kirchhoff requires 1.61×10^3 s. Therefore, from the engineering aspect, the combined acoustic monopole analysis/method of characteristics–Kirchhoff formulation is a very useful tool in preliminary design.

Based upon the prediction results of Figs. 7(a) and (b), the effect of nose shape of the train on the micro-pressure wave is shown in Fig. 13 for the three models with the same blockage ratio of 8.1%, using a combined acoustic monopole analysis/method of characteristics–Kirchhoff formulation. The model train C with the most slender nose generates the micro-pressure wave of the smallest magnitude.

VI. CONCLUDING REMARKS

Novel methods (combined acoustic monopole analysis/method of characteristics–Kirchhoff method, combined CFD–Kirchhoff method) were used to predict, the compression wave, the propagation of the wave in the tunnel, and the micro-pressure wave generated by a high-speed train. The near field flow data were obtained by acoustic monopole analysis/method of characteristics where the train is modeled by monopole sources and by three-dimensional Euler equations. For the monopole theory, second-order dipole corrections to the analytic theory were applied to render the prediction compatible with experimentally measured data. The flow data at the tunnel exit were transferred to the linear Kirchhoff formulation to predict the acoustic far field. The results, such as the entry compression wave, the pressure fluctuation in the tunnel, and the micro-pressure wave at exit, were compared with the measured data. The agreement was found to be good for both compression wave and micro-pressure wave.

The prediction methods can take into account the effect of nose shape of the train and the tunnel geometry. The control of nose shape can be an effective countermeasure to reduce the pressure fluctuation in the tunnel and the resulting micro-pressure wave. The study can be a very useful design tool in minimizing the pressure fluctuation and the booming noise. From the engineering viewpoint, the combined acoustic monopole/method of characteristics–Kirchhoff method is very efficient in that less computation time is involved. The method also provides the means of treating a long tunnel with concrete slab or ballasted tracks.

ACKNOWLEDGMENT

This work was supported by the Brain Korea 21 Project in 2001. The authors also appreciate the financial support from the government of Korea. This is a partial result of ''Aerodynamic Design and Analysis of High-Speed Railway System.''

[1] S. Ozawa, ''Studies of Micro-pressure Wave Radiated from a Tunnel Exit,'' Report of Railway Technical Research Institute, No. 1121 (1979) (in Japanese).
[2] M. Iida, T. Matsumura, K. Nakatani, T. Fukuda, and T. Maeda, ''Optimum Nose Shape for Reducing Tunnel Sonic Boom,'' ImechE paper C514/015/96 (1996).
[3] T. Maeda, T. Matsumura, M. Iida, K. Nakatani, and K. Uchida, K, ''Effect of shape train nose on compression wave generated by train entering tunnel'' in Proceedings of International Conference on Speedup Technology for Railway and Maglev Vehicles, Yokohama, Japan, 22–26 November 1993, pp. 315–319.
[4] M. S. Howe, ''Mach Number Dependence of the Compression wave generated by a High-Speed Train Entering a Tunnel,'' J. Sound Vib. **212**, 23–36 (1998).
[5] M. S. Howe, ''On Rayleigh's Computation of the 'end-correction', with an Application to the Compression Wave Generated by a Train Entering a Tunnel,'' J. Fluid Mech. **385**, 63–78 (1999).
[6] M. S. Howe, ''On the Compression Wave Generated when a High-Speed Train enters a Tunnel with a Flared Portal,'' J. Fluids Struct. **13**, 481–498 (1999).
[7] M. Iida, T. Matumura, K. Nakatani, T. Fukuda, and T. Maeda, ''Effective Nose Shape for Reducing Tunnel Sonic Boom,'' Quarter Report of Railway Technical Research Institute Vol. 38, No. 4 (1997) (in Japanese).
[8] T. Maeda, T. Matsumura, M. Iida, K. Nakatani, and K. Uchida, ''Effect of Shape of Train Nose on Compression Wave Generated by Train Entering Tunnel,'' in Proceedings of the International Conference on Speedup Technology for Railway and Maglev Vehicles, Yokohama, Japan (1993), pp. 315–319.
[9] T. Ogawa and K. Fujii, ''Numerical Simulation of Compressible Flows induced by a Train Moving into a Tunnel,'' Comput. Fluid Dynam. J. **3**, 63–82 (1994).
[10] T. Fukuda, M. Iida, K. Maeno, and H. Honma, ''Distortion of Compression Wave Propagating Through Slab Track Tunnel with Short Side Branches of Shinkansen,'' Proceedings of the 22nd International Symposium on Shock Waves, Imperial College, London, UK (1999), Paper 3040, pp. 1–6.
[11] T. Aoki, S. Nakano, K. Matuo, H. Kashimura, and T. Yasunobu, ''Attenuation and Distortion of Compression Wave and Shock Waves Propagating along High Speed Railway Model Tunnels,'' in Proceedings of the 21st International Symposium on Shock waves, Great Keppel Island, Australia (1997), Paper 1140, pp. 1–6.
[12] S. Ozawa, T. Uchida, and T. Maeda, ''Reduction of Micro-pressure Wave Radiated from Tunnel Exit by Hood at Tunnel Entrance,'' Quarter Report of Railway Technical Research Institute, Vol. 19, No. 2, pp. 77–83 (1978).
[13] S. Ozawa, T. Maeda, T. Matsumura, K. Uchida, H. Kajiyama, and K. Kanemoto, ''Countermeasures to Reduce Micro-pressure Waves Radiating from Exits of Shinkansen Tunnels,'' in *Aerodynamics and Ventilation of Vehicle Tunnels* (Elsevier Science, Oxford, 1991), pp. 253–266.
[14] T. Yoon, S. Lee, J. Hwang, and D. Lee, ''Prediction and Validation on the Sonic Boom by a High Speed Train Entering a Tunnel,'' J. Sound Vib. **247**, 195–211 (2001).
[15] S. Ozawa, T. Maeda, T. Matsumura, and K. Uchida, ''Effect of Ballast on Pressure Wave Propagating through Tunnel—1st Report; Investigation of Actual Conditions and Formulation of Basic Equations by One-dimensional Model,'' Report of Railway Technical Research Institute, Vol. 5, No. 12, pp. 9–15 (1991) (in Japanese).
[16] P. L. Roe, ''Approximate Riemann Solvers, Parameter Vectors and Difference Schemes,'' J. Comput. Phys. **43**, 357–372 (1981).
[17] A. Jameson and S. Yoon, ''Lower-Upper Implicit Schemes with Multiple Grids for the Euler Equations,'' AIAA J. **25**, 929–935 (1987).
[18] K. Fujii, ''Unified Zonal Method Based on the Fortified Solution Algorithm,'' ISAS Report No. 648, December (1992).
[19] E. Mestreau, R. Lohner, and S. Aita, ''TGV Tunnel Entry Simulations Using a Finite Element Code with Automatic Remeshing,'' AIAA Pap., 93–0890 (1993).
[20] J. L. Steger, F. C. Dougherty, and J. A. Benek, ''A Chimera Grid Scheme,'' in *Advances in Grid Generation, FED Vol. 5*, edited by K. N. Ghia (ASME, New York, 1983), pp. 59–69.
[21] M. M. Rai and K. Hessenius, ''Three Dimensional Conservative Euler Computations Using Patched Grid System and Explicit Methods,'' AIAA Pap., 86–1081 (1986).
[22] B. B. Bakerand and E. T. Copson, *The Mathematical Theory of Huygens' Principle* (Clarendon, Oxford, 1953).
[23] A. S. Lyrintzis, ''Review: The Use of Kirchhoff's Method in Computational Aeroacoustics,'' ASME J. Fluids Eng. **116**, 665–676 (1994).
[24] S. Lee, J. Kim, Y. H. Yu, and M. P. Isom, ''Prediction of Rotor High-Speed Impulsive Noise with a Combined CFD-Kirchhoff Method,'' J. Sound Vib. **207**(4), 453–464 (1997).
[25] Y. Xue and A. S. Lyrintzis, ''The Use of Rotating Kirchhoff formulation for 3-d Transonic BVI Far-field Noise,'' in 49th Annual Forum of the America Helicopter Society, St. Louis, MO (1993).
[26] C. M. Harris, *Handbook of Acoustical Measurements and Noise Control* (McGraw–Hill, New York, 1991).
[27] M. S. Howe, M. Iida, T. Fukuda, and T. Maeda, ''Theoretical and Experimental Investigation of the Compression Wave Generated by a Train Entering a Tunnel with a Flared Portal,'' J. Fluid Mech. **425**, 111–132 (2000).

Evaluation of loudness-level weightings for assessing the annoyance of environmental noise

Paul D. Schomer[a)]
US Army Construction Engineering Research Laboratory, P.O. Box 9005, Champaign, Illinois 61821

Yoiti Suzuki and Fumitaka Saito
Research Institute of Electrical Communications, Tohokku University 2-1-1 Katahira Aoba-ku, Sendai 980-8577, Japan

(Received 24 May 2000; revised 9 July 2001; accepted 15 July 2001)

Assessment of the annoyance of combined noise environments has been the subject of much research and debate. Currently, most countries use some form of the A-weighted equivalent level (ALEQ) to assess the annoyance of most noises. It provides a constant filter that is independent of sound level. Schomer [Acust. Acta Acust. **86**(1), 49–61 (2000)] suggested the use of the equal loudness-level contours (ISO 226, 1987) as a dynamic filter that changes with both sound level and frequency. He showed that loudness-level-weighted sound-exposure level (LLSEL) and loudness-level-weighted equivalent level (LL-LEQ) can be used to assess the annoyance of environmental noise. Compared with A-weighting, loudness-level weighting better orders and assesses transportation noise sources, sounds with strong low-frequency content and, with the addition of a 12-dB adjustment, it better orders and assesses highly impulsive sounds *vis-à-vis* transportation sounds. This paper compares the LLSEL method with two methods based on loudness calculations using ISO 532b (1975). It shows that in terms of correlation with subjective judgments of annoyance—not loudness—the LLSEL formulation performs much better than do the loudness calculations. This result is true across a range of sources that includes aircraft, helicopters, motor vehicles, trains, and impulsive sources. It also is true within several of the sources separately. [DOI: 10.1121/1.1402116]

PACS numbers: 43.50.Rq, 43.50.Ba [MRS]

I. INTRODUCTION

There is a need to be able to assess a noise environment that results from a variety of sources such as aircraft, roadway, train, and industrial sources. Currently, the European Community (EC) is working to develop a common methodology to assess the annoyance of combined noise environments (EC, 1996) so that noise maps can be created, initially, for every urban area with a population that exceeds 250 000. The revision of International Organization for Standardization (ISO) 1996 (ISO, April 1987) and the recent work of American National Standards Institute (ANSI) (ANSI, 1996) both attempt to deal with this topic. In addition to the noise sources listed above, sources with special sound character (impulsive, tonal, strong low-frequency content) must be able to combined into an overall noise assessment.

Currently, most countries use some form of the A-weighted equivalent level (ALEQ) to assess the annoyance of most noises. In the US we use the A-weighted day–night sound level (ADNL) and the EC plans to use A-weighted day–evening–night sound level (ADEN) for their comprehensive urban noise mapping. These metrics are used because they work fairly well. Originally Schultz (1978) suggested that all transportation noise sources could be assessed using ALEQ and ADNL. More recently, several meta-analyses of attitudinal survey data suggest that there may be differences between the community response to aircraft, train, or roadway noise for the same DNL level. Fidell *et al.* (1989) and Finegold *et al.* (1994) provide an indication that aircraft noise is more annoying than road-traffic noise at high DNLs. And, Miedema (1998) finds that aircraft noise is about 5–10 dB more annoying than road-traffic noise at all DNLs and that train noise is less annoying than road-traffic noise. Miedema's difference between train and road-traffic noise is zero at low DNLs and steadily increases with level.

In principle, the LEQ or DNL for any transportation noise source is calculated from the A-weighted sound-exposure levels of the individual vehicle pass-bys. Each train pass-by, each truck pass-by, and each plane flyby contributes some amount of A-weighted sound energy to the time-period total. The A-weighting curve, derived from the Fletcher–Munson 40-phon curve, is very roughly the inverse of the 40-phon equal-loudness-level curve. It does a pretty good job overall as a noise annoyance assessment tool, but the A-weighting filter varies only with frequency—A-weighting does not change with sound-pressure level (SPL). In contrast, the equal-loudness-level contours are a family of curves that varies with amplitude and frequency (Fig. 1). Recently, Schomer (2000) has shown that one can replace the A-weighting filter with the equal-loudness-level contours from ISO 226 (ISO, May 1987) as a dynamic filter that changes both with SPL and with frequency. Schomer shows that the use of loudness-level weighting provides for much

[a)]Author to whom correspondence should be addressed. Present address: Schomer & Associates, Inc., 2117 Robert Drive, Champaign, IL 61821; electronic mail: schomer@Schomer and associates.com

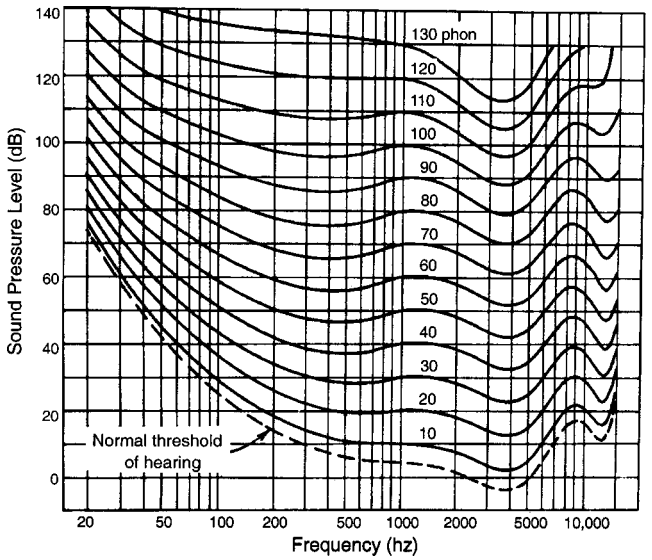


FIG. 1. Equal loudness level contours in phons from ISO 226.

better correlation with subjective annoyance responses than does the A-weighting.

In the reviews of the Schomer (2000) paper and at presentations such as Internoise and ASA meetings, one recurring question was ''Why use the curves of ISO 226? Why didn't you use loudness calculations such as ISO 532b?'' Indeed, some experimenters such as Kuwano *et al.* (1989) have shown that loudness calculations can provide better rankings across sounds than does ASEL. Suzuki *et al.* (1988) evaluated the influence of total components on both loudness and noisiness and showed that ISO 532b performed the best, even for noisiness, among several evaluation methods including PN dB. Takeshima *et al.* (1991) conducted experiments that included a multidimensional evaluation of steady noise. It was found that ''noisiness'' could be represented well by just two dimensions: *loudness* and *harshness/unpleasantness*. ''Annoyance'' was similarly well represented by these two dimensions. The results also showed that the *loudness* dimension was highly correlated with ISO 532b and thus, ISO 532b was significantly correlated with ''annoyance.'' In fact, ISO 532b correlated better with ''annoyance'' than did PN dB. Thus, it was natural that one coauthor of this paper predicted that loudness calculations using ISO 532b would work better than calculations using the equal-loudness-level contours (ISO 226–1987) to assess noise annoyance. This coauthor used all the Schomer (2000) data to perform new calculations using the method of ISO 532b (ISO, 1975) to calculate the loudness in sones and in corresponding phons. In this paper, these measures are evaluated for their ability to correlate with human assessments of annoyance in comparison with the loudness-level weighting using ISO 226.

The reader should note that this paper refers exclusively to annoyance responses. No comparison is made on the ability of these measures to assess the loudness of sounds. ISO 532b-1975 has been used to form the loudness calculations because it is the current International Standard Method.

TABLE I. Calculation of corrections inherent in the use of loudness-level-weighted sound exposure level (LLSEL)—after Schomer (2000). The entries in this table represent the averages of calculations performed separately on each event. The M-16 and G-3 events represent 60-shot volleys and the 25- and 35-mm events represent 5-shot volleys. Details on the calculations for each event can be found in Schomer (2000).

Location	Source	Number of events	ASEL (dB)	Loudness-level-weighted LLSE (dB)	LLSEL minus ASEL (dB)	Inherent correction *re* 5.0 dB	Inherent correction *re* 5.0 dB—+12 dB for gunfire
Munster	35-mm gun	5	75.0	82.8	7.8	2.8	14.8
APG	25-mm gun	10	72.4	81.2	8.8	3.8	15.8
APG	M-16 rifle	10	75.7	80.8	5.1	0.1	12.1
Munster	G-3 rifle	4	77.2	78.4	1.2	−3.8	8.2
APG	Low helicopter	7	87.1	97.9	10.8	5.8	5.8
APG	High helicopter	2	81.6	91.0	9.5	4.5	4.5
Seattle	Airplane—takeoff	24	87.6	95.5	7.9	2.9	2.9
Seattle	Airplane—landing	11	82.1	88.0	6.0	1.0	1.0
Munster	Vehicles	15	82.3	88.0	5.7	0.7	0.7
Champaign streets	Street vehicles	6	78.4	85.0	6.6	1.6	1.6
APG	V1, V2, and V3	12	90.5	96.6	6.1	1.1	1.1
APG	V4	3	79.9	87.1	7.1	2.1	2.1
Tacoma	Freeway (noisy road surface)	10	85.9	91.3	5.5	0.5	0.5
Tacoma	Freeway (noisy road surface, noisy trucks)	7	88.7	96.7	8.0	3.0	3.0
Champaign	Freeway—trucks	3	83.0	87.0	4.0	−1.0	−1.0
Champaign	Freeway—trucks	20(4)	77.7	80.5	2.8	−2.2	−2.2
Champaign	Freeway—autos	6(2)	73.9	75.9	1.9	−3.1	−3.1
APG	Train—electric—fast	2	84.1	88.6	4.5	−0.5	−0.5
APG	Train—electric—slowing	3	75.0	81.5	6.6	1.6	1.6
APG	Train—diesel—slow	2	75.9	83.8	8.0	3.0	3.0

TABLE II. Subjective responses by noise source category.

Location	Source	Subjective response *re* road traffic (dB) Base case	Alternate 1	Alternate 2
Munster	35-mm gun	16	16	16
APG	25-mm gun	16	16	16
APG	M-16 rifle	12	12	12
Munster	G-3 rifle	8	8	8
APG	Low helicopter	6	12	7
APG	High helicopter	6	12	7
Seattle	Airplane—takeoff	4	10	5
Seattle	Airplane—landing	4	10	5
Munster	Vehicles	0	0	0
Champaign	street vehicles	0	0	0
APG	V1, V2, and V3	0	0	0
APG	V4	0	0	0
Tacoma	Freeway (noisy road surface)	0	0	0
Tacoma	Freeway (noisy road surface, noisy trucks)	4	5	5
Champaign	Freeway—trucks	0	0	0
Champaign	Freeway—trucks	0	0	0
Champaign	Freeway—autos	0	0	0
APG	Train—electric—fast	−1	−5	−5
APG	Train—electric—slowing	0	−5	−5
APG	Train—diesel—slow	4	5	5

Use of any newer method to calculate loudness remains ambiguous.

II. A REVIEW OF THE LOUDNESS-LEVEL-WEIGHTING METHODOLOGY

ISO 226-1987 provides the equal-loudness-level contours (Fig. 1) in functional form for one-third-octave-band center frequencies from 20 to 12 500 Hz. Each one-third-octave-band sound-pressure level (SPL) is assigned the phon level that corresponds to that frequency and level.

Human hearing is such that short-duration sounds are not perceived to be as loud as long-duration sounds. To be perceived with full loudness, sound must be present for a duration that is longer than the integration time of the ear. In the Schomer procedure, the one-third-octave-band spectra are *fast* time weighted and sampled every 100 ms. The *fast* time weighting is used to approximate the integration time of the ear, which data indicate lies between 25 and 250 ms (e.g., Boone, 1973; Miller, 1948; Zwicker, 1966). Since the time constant for *fast* time weighting is 125 ms, 100 ms is an adequate rate with which to sample the spectra.

The time series of 100 ms, one-third-octave-band spectra is used to calculate the overall time-and frequency-summed phon level, L_L, which is given by

$$L_L = 10 \log \left\langle \sum_j \sum_i 10^{(L_{Lij}/10)} \right\rangle, \tag{1}$$

where L_{Lii} is the phon level corresponding to the ith one-third-octave band during the jth time sample.

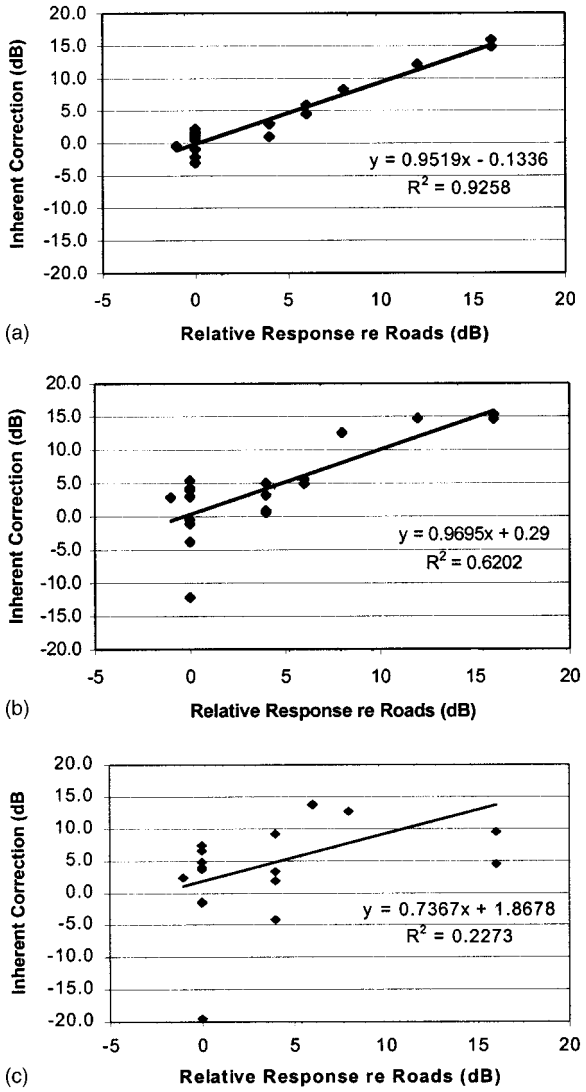


FIG. 2. Correlation between subjective judgments and the calculation indicated.

The quantity calculated by (1), L_L, has been designated as the loudness-level-weighted sound-exposure level (LLSEL). It is similar to the A-weighted sound-exposure level (ASEL) except that instead of using a filter (A-weighting) that varies only with frequency, LLSEL uses a dynamic filter that varies with both SPL and frequency. Similarly, one can calculate the loudness-level-weighted equivalent level (LL-LEQ) or loudness-level-weighted day–night level (LL-DNL).

III. DATA ANALYSIS

A. Equal-loudness-level contour data and corresponding community response data

Schomer calculated the LLSEL for a variety of sources and conditions and compared these with the known human response to these sounds. Specifically, Schomer evaluated road traffic at a variety of sites and in a variety of situations, aircraft taking off and landing, helicopter flybys, trains, and four sources of gunfire. These were compared with community responses garnered from a variety of sources.

TABLE III. Calculation of corrections inherent in the use of ISO 532b loudness-level (in phons) exposure level (LP532bEL).

Location	Source	ASEL	Loudness-weighted (ISO 532b) LP532bEL (dB)	LP532b EL minus ASEL(dB)	Inherent correction *re* 14.5 dB	Inherent correction *re* 14.5 dB—+12 dB for gunfire
Munster	35-mm gun	75.0	92.1	17.1	2.6	14.6
APG	25-mm gun	72.4	90.2	17.8	3.3	15.3
APG	M-16 rifle	75.7	92.9	17.2	2.7	14.7
Munster	G-3 rifle	77.2	92.3	15.1	0.6	12.6
APG	Low helicopter	87.1	106.6	19.5	5.0	5.0
APG	High helicopter	81.6	101.7	20.1	5.6	5.6
Seattle	Airplane—takeoff	87.6	103.1	15.4	0.9	0.9
Seattle	Airplane—landing	82.1	97.2	15.1	0.6	0.6
Munster	Vehicles	82.3	100.8	18.5	4.0	4.0
Champaign	Street vehicles	78.4	97.2	18.8	4.3	4.3
APG	V1, V2, and V3	90.5	103.9	13.4	−1.1	−1.1
APG	V4	79.9	94.0	14.0	−0.5	−0.5
Tacoma	Freeway (noisy road surface)	85.9	103.4	17.5	3.0	3.0
Tacoma	Freeway (noisy road surface, noisy trucks)	88.7	106.5	17.8	3.3	3.3
Champaign	Freeway—trucks	83.0	100.6	17.6	3.1	3.1
Champaign	Freeway—trucks	77.7	88.4	10.7	−3.8	−3.8
Champaign	Freeway—autos	73.9	76.3	2.4	−12.1	−12.1
APG	Train—electric—fast	84.1	101.5	17.4	2.9	2.9
APG	Train—electric—slowing	75.0	94.9	19.9	5.4	5.4
APG	Train—diesel—slow	75.9	95.4	19.5	5.0	5.0

In his analysis, the physical parameter of interest is the difference between the LLSEL and the ASEL for various noise events. For example, the literature suggests that aircraft noise is perhaps 5 dB more annoying than road-traffic noise for the same A-weighted sound level. Suppose that the difference, Δ_R, between LLSEL and ASEL for motor vehicles is, on average, 10 dB, and that the Δ_A for aircraft flybys is, on average, 15 dB. Then, the 5-dB adjustment that is re-

TABLE IV. Calculation of corrections inherent in the use of ISO 532b loudness (in sones—converted to decibels) exposure level (LS532bEL).

Location	Source	ASEL	Sone-seconds (ISO 532b) N-s	10 log(N-s) LS532bEL (dB)	LS532bEL minus ASEL (dB)	Inherent correction *re* 76.0 dB	Inherent correction *re* 76.0 dB—+12 dB for gunfire
Munster	35-mm gun	75.0	1 501.0	143.4	68.4	−7.6	4.4
APG	25-mm gun	72.4	1 582.0	145.8	73.5	−2.5	9.5
APG	M-16 rifle	75.7	4 448.5	160.5	84.8	8.8	20.8
Munster	G-3 rifle	772	2 706.9	153.8	76.7	0.7	12.7
APG	Low helicopter	87.1	15 225.7	176.8	89.7	13.7	13.7
APG	High helicopter	81.6	9 044.9	171.4	89.8	13.8	13.8
Seattle	Airplane—takeoff	87.6	7 095.1	167.0	79.3	3.3	3.3
Seattle	Airplane—landing	82.1	4 253.6	159.9	77.9	1.9	1.9
Munster	Vehicles	82.3	7 145.9	165.6	83.3	7.3	7.3
Champaign	Street vehicles	78.4	4 149.2	159.1	80.7	4.7	4.7
APG	V2, V2, and V3	90.5	10 290.5	173.1	82.6	6.6	6.6
APG	V4	79.9	4 879.0	159.8	79.9	3.9	3.9
Tacoma	Freeway (noisy road surface)	85.9	4 765.4	160.3	74.4	−1.6	−1.6
Tacoma	Freeway (noisy road surface, noisy trucks)	88.7	4 882.8	160.5	71.8	−4.2	−4.2
Champaign	Freeway—trucks	83.0	5 172.8	163.0	79.9	3.9	3.9
Champaign	Freeway—trucks	77.7	2 924.5	152.3	74.7	−1.3	−1.3
Champaign	Freeway—autos	73.9	683.7	130.4	56.5	−19.5	−19.5
APG	Train—electric—fast	84.1	4 836.3	162.4	78.3	2.3	2.3
APG	Train—electric—slowing	75.0	2 852.2	154.6	79.6	3.6	3.6
APG	Train—diesel—slow	75.9	4 408.9	161.0	85.2	9.2	9.2

TABLE V. Summary of alternative subjective responses and corresponding calculations.

Location	Source	Subjective response *re* road traffic (dB)			LLSEL Inherent correction *re* 5.0 dB—+12 dB for gunfire	LP532bEL Inherent correction *re* 14.5 dB—+12 dB for gunfire	LS532bEL Inherent correction *re* 76.0 dB—+12 dB for gunfire
		Base case	Alternate 1	Alternate 2			
Munster	35-mm gun	16	16	16	14.8	14.6	4.4
APG	25-mm gun	16	16	16	15.8	15.3	9.5
APG	M-16 rifle	12	12	12	12.1	14.7	20.8
Munster	G-3 rifle	8	8	8	8.2	12.6	12.7
APG	Low helicopter	6	12	7	5.8	5.0	13.7
APG	High helicopter	6	12	7	4.5	5.6	13.8
Seattle	Airplane—takeoff	4	10	5	2.9	0.9	3.3
Seattle	Airplane—landing	4	10	5	1.0	0.6	1.9
Munster	Vehicles	0	0	0	0.7	4.0	7.3
Champaign	Street vehicle	0	0	0	1.6	4.3	4.7
APG	V1, V2 and V3	0	0	0	1.1	−1.1	6.6
APG	V4	0	0	0	2.1	−0.5	3.9
Tacoma	Freeway (noisy road surface)	0	0	0	0.5	3.0	−1.6
Tacoma	Freeway (noisy road surface, noisy trucks)	4	5	5	3.0	3.3	−4.2
Champaign	Freeway—trucks	0	0	0	−1.0	3.1	3.9
Champaign	Freeway—trucks	0	0	0	−2.2	−3.8	−1.3
Champaign	Freeway—autos	0	0	0	−3.1	−12.1	−19.5
APG	Train—electric—fast	−1	−5	−5	−0.5	2.9	2.3
APG	Train—electric—slowing	0	−5	−5	1.6	5.4	3.6
APG	Train—diesel—slow	4	5	5	3.0	5.0	9.2

quired when using A-weighting is automatically incorporated in the physical measurements if one uses LLSEL, because the difference between Δ_A and Δ_R is 5 dB.

Schomer used meta-analysis on attitudinal survey psychoacoustical data to describe the differences between transportation noise sources. Previously recorded motor vehicle data supplemented by new one-third-octave-band measurements on motor vehicles, trains, and airplanes were used to supply the acoustical data for transportation noise sources. Previously gathered single-event psychoacoustical judgments, along with corresponding recorded acoustical data, were used for impulsive noises. Table I gives results from Schomer's analysis of the physical data. This table includes a column indicating the difference between LLSEL and ASEL for each category. Then, for each category, the next column gives the difference relative to road traffic by subtracting the average difference for road traffic. (The data for the noisy trucks on a noisy road are not included in this average.) Finally, the last column in Table I gives the previous column with 12 dB added to the indicated values for gunfire sources. The 12 dB are added since ISO 1996 includes a penalty of 12 dB on impulsive sources.

Table II gives the relative subjective responses to the same sources as contained in Table I. In Table II, the relative response to gunfire sources compared to vehicular sources is taken from actual measured values for those particular data (Schomer and Wagner, 1995). The relative response for aircraft as compared to road traffic is taken from the Finegold *et al.* (1994) meta-analysis. Finegold *et al.* show a penalty of about 4 dB for aircraft and a bonus of at most 1 dB for trains. For example, previous work by Schomer and Wagner (1995) and by Fields and Powell (1985) shows only a small difference, typically 2 dB or less, in community response to helicopters as compared with fixed-wing aircraft. So, for helicopters, 2 dB is added to the 4-dB penalty for aircraft for a total helicopter penalty of 6 dB. Noisy trucks on a noisy road surface are arbitrarily given a penalty of 4 dB. Miedema (1998) has shown considerably larger adjustments as compared with Finegold. Therefore, as an alternate set of subjective responses, column 3 in Table II shows the ''Alternate 1'' responses. For these, the basic aircraft penalty is changed from 4 to 10 dB and the basic train bonus is changed from 1 to 5 dB. Helicopters are given a penalty of 12 dB, and the penalty for noisy trucks on a noisy surface is raised to 5 dB. As a second alternate, the currently proposed penalties in a

TABLE VI. Linear-fit equation parameters and *R*-squared values for each response alternative and each calculation method.

Case	Calculation	$Y=Kx+B$		R^2
		K	B	
Base case	LLSEL	0.9519	−0.1336	0.9258
	LP532bEL	0.9695	0.2900	0.6202
	LS532bEL	0.7367	1.8678	0.2273
Alternate 1	LLSEL	0.6404	0.5526	0.6514
	LP532bEL	0.5905	1.2853	0.3576
	LS532bEL	0.5842	1.9738	0.2222
Alternate 2	LLSEL	0.8005	0.5844	0.8160
	LP532bEL	0.7693	1.1963	0.4867
	LS532bEL	0.6273	2.3941	0.2054

TABLE VII. Calculated corrections versus subjective response to gunfire sounds.

Location	Source	Subjective response *re* road traffic	LLSEL Inherent correction *re* 5.0 dB—+12 dB for gunfire	LP532bEL Inherent correction *re* 14.5 dB—+12 dB for gunfire	LS532bEL Inherent correction *re* 76.0 dB—+12 dB for gunfire
Munster	35 mm gun	16	14.8	14.6	4.4
APG	25 mm gun	16	15.8	15.3	9.5
APG	M-16 rifle	12	12.1	14.7	20.8
Munster	G-3 rifle	8	8.2	12.6	12.7

draft revision to ISO 1996 are used. For this alternate, the aircraft penalty is 5 dB, the helicopter penalty is 7 dB, the train bonus is 5 dB, and, again, the penalty for noisy trucks on a noisy road surface is arbitrarily set to 5 dB.

B. Calculations using the loudness methods of ISO 532b

For the above data analysis using the equal-loudness-level contours, a fast-time-weighted one-third-octave-band spectrum was sampled every 100 ms for each event. Table I lists the number of events that each category represents. For the ISO 532b analysis, the loudness in sones and phons has been calculated separately for each event using the ISO standard for each 100-ms one-third-octave-band spectrum. Just as environmental noise annoyance is currently assessed using ASEL and ALEQ, and just as Schomer (2000) calculated LL-SEL, analogous calculations are performed herein using ISO 532b calculated sones and phones. For the phon analysis, the 100-ms phon samples for each event have been summed on an energy basis to form the ISO 532b loudness (in phons) exposure level (LP532bEL). Table III contains the results for the analysis using LP532bEL.

For the sone analysis, the sones for every 100-ms sample for each event have been summed to form the total ''sone seconds'' for that event. These total-sone-second values have been converted to a decibel-like number by taking 10 times the logarithm base 10 of the total-sone-seconds. These are represented herein by the symbol LS532bEL. As a further variant, the decibel conversion was performed using 10 times the logarithm base 2. However, the correlation between the base 2 and base 10 conversions is so high that only the base 10 conversion is further considered. Table IV contains results when LS532bEL is used.

IV. DISCUSSION

A. Basic results

Table V summarizes the results from Tables I through IV. It contains the community response adjustments relative to road-traffic noise (including the two alternatives as discussed above), and the physical data for the difference between the measure indicated and A-weighting relative to road-traffic noise and including, in each case, the 12-dB adjustment for gunfire noises. Figures 2(a), (b), and (c) plot the measures in Table V compared with the base case response values for LLSEL, LP532bEL, and LS532bEL, respectively. Each figure contains an equation fit to the line and its *R*-squared value. Clearly, LLSEL does the best job of correlating with these community response judgments.

However, one may suggest that the above analyses depend on the response judgments chosen. Therefore, Table VI contains the linear-fit equation parameters and *R*-squared values for each response alternative and each calculation method. For each case, LLSEL correlates much better with the annoyance response data than does either of the methods based on ISO 532b.

B. Alternative comparisons

In the above analysis, all the data are compared and evaluated as a group. In the following analysis, individual categories of sound are evaluated separately.

1. Gunfire data

Table VII contains just a summary of the gunfire data results. Clearly, only the LLSEL data correlate well with the field-measured subjective responses. The LP532bEL measure correctly orders the data but does not reflect the magni-

TABLE VIII. Calculated corrections versus subjective response to aircraft sounds.

Location	Source	Subjective response *re* road traffic (dB)			LLSEL Inherent correction *re* 5.0 dB	LP532bEL Inherent correction *re* 14.5 dB	LS532bEL Inherent correction *re* 76.0 dB
		Base case	Alternate 1	Alternate 2			
APG	Low helicopter	6	12	7	5.8	5.0	13.7
APG	High helicopter	6	12	7	4.5	5.6	13.8
Seattle	Airplane—takeoff	4	10	5	2.9	0.9	3.3
Seattle	Airplane—landing	4	10	5	1.0	0.6	1.9

TABLE IX. Calculated corrections versus subjective responses to train sounds.

Location	Source	Subjective response *re* road traffic (dB) Base case	Alternate 1	Alternate 2	LLSEL Inherent correction *re* 5.0 dB	LP532bEL Inherent correction *re* 14.5 dB	LS532bEL Inherent correction *re* 76.0 dB
APG	Train—electric—fast	−1	−5	−5	−0.5	2.9	2.3
APG	Train—electric—slowing	0	−5	−5	1.6	5.4	3.6
APG	Train—diesel—slow	4	5	5	3.0	5.0	9.2

tude of the differences from one to another. The LS532bEL measure does not correctly order the data or reflect the field-measured subjective response differences.

2. Aircraft data

Table VIII contains just a summary of the aircraft data. Laboratory and survey data reveal only about a 2-dB difference, at most, between fixed-wing aircraft and helicopters. Clearly, the LLSEL data correlate in size and direction with this difference. The LP532bEL measure yields a believable difference of 5 dB, although this difference is larger than expected. The LS532bEL measure yields a systematic difference of 10 dB, which is clearly too large.

3. Train data

Table IX contains just the train data. Note that these are for American trains running on less-than-perfect tracks. Based only on these train data, each of the three measures is plausible. However, the LP532bEL procedure rates the slowing electric train noise as about equal to the diesel train noise. This result is perhaps less probable.

4. Road-traffic data

Table X contains all of the road-traffic data. Several factors can be observed from these data. First, consider just the urban street traffic data. Using LLSEL the offset is quite constant, ranging from 5.7 to 7.1 dB. The measures based on ISO 532b show corresponding ranges of 3 to 4 dB and are certainly plausible.

Second, consider just the highway data. Except for the noisy trucks on a noisy surface, the LLSEL range is from 1.9 to 5.5 dB. And, automobiles on a quiet road surface yield the smallest value (1.9 dB), and trucks on a noisy surface yield the largest value (5.5 dB). The noisy trucks on the noisy road yield a value of 8 dB. In contrast, the LP532bEL measure yields a range of 2 to 17 dB, and the noisy trucks are not rated more poorly than the regular trucks. Similarly, the LS532bEL measure yields a range of 13 dB and the noisy trucks are rated 3 dB better than corresponding regular trucks on the same road surface at the same speed.

Third, consider the highway data compared with the urban traffic data. Using LLSEL, the urban traffic is rated a little worse than the highway traffic. This probably fits reality, since nearly all of the urban sounds were for vehicles accelerating, whereas the highway data are for vehicles at constant speed. The LP532bEL and LS532bEL measures do not show this distinction between the urban and highway situations.

V. CONCLUSIONS

Loudness-level-weighted sound-exposure level (LLSEL) and loudness-level-weighted equivalent level (LL-LEQ) can be used to assess the annoyance of environmental noise. Compared with A-weighting, loudness-level weighting better orders and assesses transportation noise sources, and with the addition of a 12-dB adjustment, loudness-level weighting better orders and assesses highly impulsive sounds.

The LLSEL does a significantly better job of assessing the annoyance of combined noise sources as compared with the methods based on loudness calculations using ISO 532b, examined herein. For any set of response judgments, the correlation with LLSEL is always much higher. Further, for just

TABLE X. Calculated corrections versus subjective responses to vehicle sounds.

Location	Source	Subjective response *re* road traffic	LLSEL Inherent correction *re* 5.0 dB	LP532bEL Inherent correction *re* 14.5 dB	LS532bEL Inherent correction *re* 76.0 dB
Munster	Vehicles	0	0.7	4.0	7.3
Champaign	Street vehicles	0	1.6	4.3	4.7
APG	V1, V2, and V3	0	1.1	−1.1	6.6
APG	V4	0	2.1	−0.5	3.9
Tacoma	Freeway (noisy road surface)	0	0.5	3.0	−1.6
Tacoma	Freeway (noisy road surface, noisy trucks)	4–5	3.0	3.3	−4.2
Champaign	Freeway—trucks	0	−1.0	3.1	3.9
Champaign	Freeway—trucks	0	−2.2	−3.8	−1.3
Champaign	Freeway—autos	0	−3.1	−12.1	−19.5

assessing gunfire or highway traffic alone, there is a clear benefit to the use of LLSEL over the methods based on the loudness calculations using ISO 532b. For trains or urban road traffic alone, no method is clearly better. Across urban and highway traffic together, LLSEL also appears to be superior. For aircraft, the LLSEL and the LP532bEL measures are better than the LS532bEL measure.

Since type 1, hand-held one-third-octave-band instruments are readily available at relatively low costs, it would be inexpensive to implement LLSEL and LL-LEQ capabilities in these hand-held instruments. Thus, significant improvements can be made to the measurement and assessment of environmental noise without resorting to the large number of adjustments that are required when assessing sound using just the A-weighting.

ANSI (**1996**). ANSI S12.9-1996—Part 4, ''American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound—Part 4: Noise Assessment and Prediction of Long-term Community Response'' (Acoustical Society of America, New York).

Boone, M. M. (**1973**). ''Loudness measurements on pure tone and broadband impulsive sounds,'' Acustica **29**, 198–204.

EC (**1996**). ''Future Noise Policy,'' European Commission Green Paper, European Union, Brussels, 1996.

Fidell, J., and Barber, D. J., and Schultz, T. J. (**1989**). ''Updating a dosage-effect relationship for the prevalence of annoyance due to general transportation noise,'' Technical Rpt. HSD-TR-89-009, Human System Division, US Air Force Systems Command, Brooks Air Force Base, TX.

Fields, J. M., and Powell, C. A. (**1985**). ''A Community Survey of Helicopter Noise Annoyance Conducted Under Controlled Noise Exposure Conditions,'' NASA TM-86400, March 1985.

Finegold, L. S., Harris, S. C., and von Gierke, H. E. (**1994**). ''Community annoyance and sleep disturbance: Updated criteria for assessing the impacts of general transportation noise on people,'' Noise Control Eng. J. **42**(1), 25–30.

ISO (**1975**). ''Acoustics—Method for calculating loudness level,'' ISO Recommendation R 532b, International Organization for Standardization (ISO), Geneva, Switzerland, December 1966.

ISO (**April 1987**). ''Acoustics—Description and measurement of environmental noise—Part 2: Acquisition of data pertinent to land use,'' ISO 1996/2-1987, International Organization for Standardization (ISO), Geneva, Switzerland, 15 April 1987.

ISO (**May 1987**). ''Acoustics—Normal equal-loudness level contours,'' ISO 226, International Organization for Standardization (ISO), Geneva, Switzerland, May 1987.

Kuwano, S., Namba, S., and Miura, H. (**1989**). ''Advantages and disadvantages of A-weighted sound pressure level in relation to subjective impression of environmental noises,'' Noise Control Eng. J. **33**, 107–115.

Miedema, H. M. E. (**1998**). ''Exposure-response relationships for transportation noise,'' J. Acoust. Soc. Am. **104**, 3432–3445.

Miller, G. A. (**1948**). ''Perception of short bursts of noise,'' J. Acoust. Soc. Am. **20**.

Schomer, P. D., and Wagner, L. R. (**1995**). ''Human and community response to military sounds II. Results from field-laboratory tests of sounds of small arms, 25-mm cannons, helicopters, and blasts,'' Noise Control Eng. J. **43(1)**, 1–13.

Schomer, P. D. (**2000**). ''Loudness-level weighting for environmental noise assessment,'' Acust. Acta Acust. **86**(1), 49–61.

Schultz, T. J. (**1978**). ''Synthesis of social surveys on noise annoyance,'' J. Acoust. Soc. Am. **64**, 377–405.

Suzuki, Y., Kono, S., and Sone, T. (**1988**). ''An experimental consideration of the evaluation of environmental noise with tonal components,'' J. Sound Vib. **127**, 475–484.

Takeshima, H., Suzuki, Y., and Sone, T. (**1991**). ''Evaluation of steady noise from a multidimensional point of view,'' J. Sound Vib. **151**, 519–528.

Zwicker, E. (**1966**). ''Ein Beitrag zur Lautstarkemessung Impulshaltiger Gerausche,'' Acustica **17**, 11–22.

Mechanism of sound absorption by seated audience in halls

Noriko Nishihara[a] and Takayuki Hidaka
Takenaka Research & Development Institute, 1-5-1, Otsuka, Inzai, Chiba 270-1395, Japan

Leo L. Beranek
975 Memorial Drive, #804, Cambridge, Massachusetts 02138

(Received 16 October 2000; revised 10 August 2001; accepted 15 August 2001)

Four methods are explored for predicting the reverberation times in fully occupied halls for music as related to the sound absorption by their audiences. The methods for providing audience absorptions include two that use reverberation chambers, namely, the ISO 354 method (and other similar standards) (ISO) and Kath and Kuhl's method (K & K) [Acustica **15**, 127–131 (1965)], and two that use average data from halls, i.e., Beranek's method (COH) [*Concert and Opera Halls: How They Sound* (Acoustical Society of America, Melville, NY, 1996)], and the average audience power-per-seat absorption which in practice is multiplied by the number of seats (ΔA). These methods are applied to the calculation of reverberation times in six existing halls, fully occupied, and the results were compared with actual measurements. The COH method was best for predictions over the entire frequency range. The K & K method showed the highest accuracy at mid-frequencies. Both the ISO and the K & K methods yielded wide differences for the measurements in the 125- and 250-Hz bands. The ΔA method was as good as the COH method when the measurements for the six halls were averaged, but showed a wide spread in the predictions around the average because it does not consider the degree of upholstering of the seats. It was hypothecated by the authors that the principal reasons for the ISO and K & K discrepancies at low frequencies were (a) differences between the degree of sound diffusion in actual halls and that in reverberation chambers, and (b) lack of information on the mechanisms of absorption of sound by people seated side-by-side in rows, particularly for near-grazing incidence sound fields. First, this article explores the sound diffusivity in a reverberation chamber and in the halls using CAD models. A probability density function of the incident angles of the sound rays that impinge on the audiences is defined and was measured for each case. Using a unique method, the sound absorption coefficient of each portion of the body and chair in a seated audience was determined in an anechoic chamber as a function of the incident angle of a sound wave. With adjustments from these findings, the K & K method can be made to equal the COH method in accuracy at all frequencies. Its forte is that it can be used for the determination of the sound absorption of occupied chairs from measurements of a limited number in a reverberation chamber.
[DOI: 10.1121/1.1409970]

PACS numbers: 43.55.Br, 43.55.Fw, 43.55.Gx, 43.55.Mc [JDQ]

I. INTRODUCTION

Several studies on sound absorption by seated audiences have been published. Included are Kath and Kuhl (1964, 1965), Davis *et al.* (1994), and Bradley (1992, 1996). They examined several approaches for more precise prediction of reverberation time (RT) in halls by using the results of audience absorption measurements in reverberant chambers. In general, they have concluded that the reverberation times in an actual occupied hall cannot be predicted satisfactorily from absorption coefficients measured in a reverberation chamber. Beranek (1962, 1996) and Beranek and Hidaka (1998) proposed a method for predicting the reverberation times in occupied halls (with Sabine's formula) in which the absorption coefficients of audiences were based on measurements of reverberation times made in many halls.

This study attempts to make clear the accuracy and the limitation of each prediction method with an eye toward its use in new halls of modern shapes and/or audience chairs of new designs. In other words, it is intended (a) to search for and establish an appropriate method to predict reverberation times in occupied halls, with focus on the accuracy of the audience absorption coefficients used in the above methods, and (b) to determine the reasons for the differences in the absorption coefficients measured in reverberation chambers and those measured in actual halls.

II. COMPARISON OF RT PREDICTION METHODS

The residual absorption, which is defined as that for a room, separate from the audience, orchestra, and any strongly absorbing materials (Beranek and Hidaka, 1998), generally contributes to the total absorption power of a hall about 20% to 30% at mid- and high frequencies. Therefore, even admitting $\pm 3\%$ error in the determination of residual surface areas, such inaccuracy would not seriously affect the accuracy of predictions of RT in occupied halls. On the other hand, measurements show that of the total sound absorption

[a]Author to whom correspondence should be addressed; electronic mail: nishihara.noriko@takenaka.co.jp

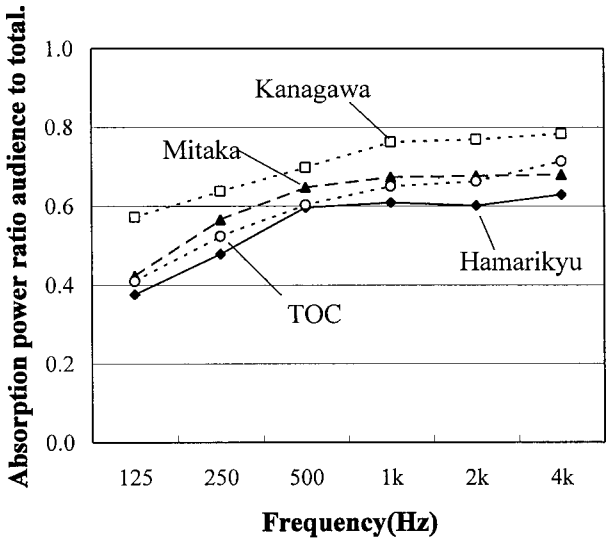


FIG. 1. Average ratio of the absorption power of the audience to the total absorption power in each of the four halls, Hamarikyu-Asahi Concert Hall, Kanagawa Auditorium, Mitaka Concert Hall, and Tokyo Opera City Concert Hall. The audience absorption power was derived from reverberation-time measurements made before and after installation of chairs (occupied). The halls were completely finished before the seats were installed and all debris was removed.

in a hall, 60% to 80% is due to the audience, so that its contribution to RT predominates (Fig. 1). Hence, the remaining contribution is caused by air absorption, which is only important at high frequencies.

Audience absorption coefficients are usually determined from either (a) measurements made in a reverberation chamber involving a few occupied seats or (b) measurements made in an actual occupied hall. In this study, four methods for determining audience absorption coefficients for use in halls for music are compared. For either (a) or (b) the sound absorption power for the audience as a whole (where air absorption can be neglected or subtracted) has been determined by

$$a_{\text{occup}} = (0.161V/T_{\text{occup}}) - a_{\text{other}}, \tag{1}$$

where a_{other} is determined from the same equation as that within the parentheses but with T being that measured before the seats are installed. Each "a" is the product of an area and an average sound absorption coefficient. Obviously, the absorption power must be determined at each frequency of interest.

A. Four methods for determining sound absorption coefficients for seated audiences

1. The ISO method

The standards relating to the measurement of absorption coefficients (or powers) of the seats, occupied or unoccupied, in reverberation chambers [ISO 354 (1983), ASTM C423-90a (1990), DIN 52212 (1961), BS 3638 (1987), and JIS A 1409 (1998)] stipulate that a group of them be placed near the center of the floor, 1 m or more away from each wall, and that room diffusers or microphones be placed in somewhat the same way as they are in an actual hall. A barrier around the seating group is recommended in some cases.

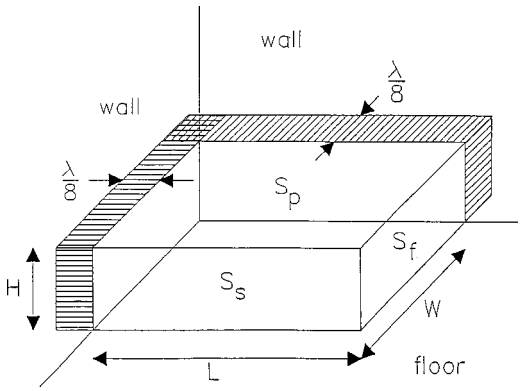


FIG. 2. Schematic seating array used in the Kath and Kuhl method. When the seating block is expressed as $L \times W \times H$, the equivalent areas of edges, S_p', S_s', and S_f', exposed to the sound field in the reverberant chamber are defined as Eq. (2) in order to compensate for the Waterhouse effect.

The reverberant chamber of Takenaka R&D, which was used for this research, has seven walls with no parallel surfaces and has a volume of 332.8 m^3 and a surface area of 289.2 m^2, in which 20 curved diffusers, each with an area of 1.8 m^2, are randomly suspended. Based on recommendations in the standards, 20 seats (seating arrays of 5 seats×4 rows with row-to-row spacing of 950 mm) chosen from those actually used in the six halls[1] were installed at the center of the chamber. During measurements, all five surfaces of these seating blocks were exposed to the sound field, and Eq. (1) is used to calculate the absorption coefficients, using as the area that area over which the seat block is located, S_p. When the absorption coefficients determined by this method are used to predict reverberation times in an actual hall, the hall area chosen is also that which lay directly beneath the seating blocks—no edge correction is applied.

2. K & K method (Kath and Kuhl, 1964, 1965)

This method assumes the seated audience to be a rectangular-parallelepiped, solid absorber, covering an area S_p. The seats were installed in a 90-degree corner of the reverberation chamber, in order to minimize the edge effect, and the top, side, and front part of the solid block were given the sound absorption coefficients α_p, α_s, and α_f, respectively (Fig. 2). As shown, λ/8-wide strips are added to the edge surfaces at the walls to compensate for the additional absorption by the Waterhouse effect, where λ is wavelength (Waterhouse, 1955). Then, the equivalent area of each edge for the calculation is expressed as follows:

$$\begin{aligned} S_p' &= S_p + (L+W)\times\lambda/8 + (\lambda/8)^2 \quad \text{(top)} \\ S_s' &= S_s + \lambda/8 \times 1.2 \text{ m} \quad \text{(side)} \\ S_f' &= S_f + \lambda/8 \times 1.2 \text{ m} \quad \text{(front)}. \end{aligned} \tag{2}$$

During the measurements, the edges of the seating array on the two sides that were exposed to the sound field were surrounded by stiff wooden barriers (1.2 m high, equal to the height of the head of a seated person). The absorption coefficients, α_p, α_s, and α_f, were determined by using simultaneous equations and removing the barrier from S_s or S_f as appropriate, for five arrangements of seating arrays (5 chairs

TABLE I. Absorption powers of audience per person obtained from survey of 20 acoustical textbooks. Their median value was used for the ΔA method.

	Frequency (Hz)					
	125	250	500	1k	2k	4k
Lord and Templeton (1996)	0.33	0.40	0.44	0.45	0.45	0.45
Furrer (1964)	0.20	0.40	0.55	0.60	0.60	0.50
Kimura (1986)	0.20	0.32	0.40	0.43	0.43	0.40
Nagata (1991)	0.25	0.34	0.41	0.43	0.42	0.41
Median	**0.23**	**0.37**	**0.43**	**0.44**	**0.44**	**0.43**

$\times 4$ rows, 4 chairs$\times 5$ rows, 3 chairs$\times 6$ rows, 6 chairs$\times 3$ rows, and staggered 5 chairs$\times 4$ rows). The wooden barrier utilized in this study is made of two layers of 12-mm-thick plywood nailed to a wooden frame and reinforced with wooden battens forming squares $40\times 40\,\mathrm{cm}^2$. From the preliminary measurement in the reverberant chamber, it was concluded that the absorption power of this barrier is less than 1 m^2, which has little influence on the results discussed in the next section.

When an absorption coefficient at a particular frequency determined by the K & K method is used to calculate the absorption power of the audience in a real hall, the areas of the exposed edges are determined accurately from the architectural drawings. The following principles for obtaining that total area are applied:

(i) For edges adjacent to balcony fronts or walls, sound absorption by them was disregarded in the calculation.
(ii) The side edges (of two seating blocks) that face each other, e.g., at an aisle, are treated as one edge as though they are combined.
(iii) The absorption by the other edges, S_s and S_f, is determined by first taking their length and multiplying it by the height (1.2 m), then multiplying that product by the appropriate absorption coefficients, α_s and α_f.

3. COH method

According to Beranek (1996), the absorption power of an audience, including that of the orchestra if present, is defined by the first term in the denominator of Sabine's formula,

$$\mathrm{RT}=\frac{0.161\cdot V}{S_T\cdot\alpha_T+S_R\cdot\alpha_R+S_N\cdot\alpha_N+4mV}, \tag{3}$$

where V denotes room volume; m is air absorption coefficient; and S is a surface area. The subscript T specifies the audience area, including an edge effect equal to the area of a strip of 0.5 m wide around the seating area, except where a side area is adjacent to a balcony front or a wall. The subscript R specifies the residual area as discussed earlier. The subscript N applies to any area of special sound-absorbing material installed for sound control—there may be several N terms if there are several different types of absorbing materials. The vertical parts of steps on floors of the seating area are not included in the area calculations.

The absorption coefficients of the audience, α_T, were determined by this method from RT measurements made in 21 halls with full occupancy (Beranek and Hidaka, 1998). The occupied seats in those halls were classified into three categories depending on the degree of seat upholstering: heavily, medium, or lightly upholstered, and their average in each category was taken as the sound absorption coefficient for that category. Then, the values of the residual absorption coefficients α_R were estimated from data in Table III of Beranek and Hidaka (1998), unless measurements were made before the seats were installed.

4. ΔA method

For this simple method the absorption power of each seated person in a hall is determined by dividing a_{occup} [Eq. (1)] by the number of seats. In practice, the absorption power per person is measured in a number of halls and the average used for the predictions of RTs in other halls. The results of this method have been reported in the literature (Furrer, 1964; Kimura, 1986; Nagata, 1991; Lord and Templeton, 1996). It is often used because the more tedious determination of the audience area from architectural drawings is unnecessary. From a literature survey of more than 20 textbooks, it was found that the values shown in Table I are usually used. Their median value was employed in this study.

B. Comparison of absorption powers of seated audiences

The total sound absorptions by seated audiences were obtained from RT measurements before and after installing the chairs (with audience) in six halls (named in Tables II and III), where the area actually covered by the seats was defined as the audience area S_p (see Fig. 2). The upholstering of the seats in those halls is light, except for Hamarikyu-Asahi, where the level of upholstering is medium.

Figure 3 compares the average values of the ratios of the measured to the calculated total absorption power of the seated audiences in the six halls. The COH method has the highest accuracy for all frequency ranges, because it is based on the average value of the sound absorption coefficients measured in 21 existing halls and the level of upholstering of the chairs is considered. The ΔA method looks good on average, but it has the largest variation range in the RT values for the different halls which is indicated by the vertical bars. This is to be expected because ΔA does not take into account the upholstering and the arrangement of the seats. The K & K method gives the best results at mid-frequencies, and is

TABLE II. Six halls in this study for which the occupied data were measured accurately. V is room volume; N is the number of seats; S is the total surface area; S_{ta} is the total area under and around the seating, i.e., aisles, rear and front and side areas; S_a is the area over which the audience sits, with 0.5-m edge effect included; S_p, S_s, and S_f are the area of the upper, side, and front part based on the principles of K & K method, respectively.

	V (m^3)	N	S (m^2)	S_{ta} (m^2)	S_a (m^2)	S_p (m^2)	S_s (m^2)	S_f (m^2)
Mitaka Concert Hall	5500	625	2315	451	413	312	88	32
Hamarikyu Asahi Hall	5800	552	2570	412	395	277	76	26
New National Theater Medium Theater	7200	1038	2786	834	614	534	116	50
New National Theater Opera House	14 500	1810	5666	1411	1153	891	240	40
Vienna, Musikvereinssaal	15 000	1573	4102	1059	955	690	198	78
Tokyo Opera City Concert Hall	15 300	1632	5843	1312	1052	793	261	39

fairly good at frequencies above 500 Hz. However, a large deviation, about a factor of 1.5, is observed for the 125- and 250-Hz frequency bands. The K & K method is superior to the ISO method at medium and high frequencies.

From these results it is seen that of the four, the COH method is the best for predicting the reverberation times in halls for music if the degree of upholstering on chairs of conventional design is known. But what about halls in which it is hoped to use new designs of chairs? Because the number of sample chairs is usually small, their absorption properties will have to be measured in reverberation chambers. The K & K method is obviously more effective than the ISO method for frequencies of 500 Hz and above. To be fully effective, the discrepancies for the low-frequency bands must be solved. In the remainder of this article, we shall attempt to diagnose the reason for this low-frequency discrepancy and to recommend means for eliminating it.

III. PHYSICAL DIFFUSION IN A REVERBERANT CHAMBER AND IN A HALL

In regard to the diffusivity of the sound field in a reverberation chamber, Balachandran *et al.* (1967/1968) reported that it is almost perfect for mid- and high frequencies if sound diffusing panels are installed. Dämming (1991) has found that such panels are often not effective at low frequencies. It is generally believed that the diffusivity of the sound field in a hall for music is imperfect at all frequencies, although there is little published evidence to support this conclusion. At least, it seems reasonable to assume that ideal diffusion does not occur in the low frequency range, particularly when the hall is occupied and most of the sound absorption takes place on horizontal surfaces. It must be remarked that this discussion applies to halls with reverberation times greater than 1.5 s, at mid-frequencies when fully occupied. The absorption coefficient of a seated audience measured in a reverberation chamber, α_s, will equal the statistical (theoretical) absorption coefficient, α_{stat}, if the diffusion of the sound field is perfect and if excess absorption by the edge effect and other disturbing factors (Makita and Hidaka, 1987) can be excluded.

$$\alpha_s \cong \alpha_{\text{stat}} = \frac{1}{\frac{1}{2}} \int_0^{\pi/2} \alpha(\theta) \cos\theta \sin\theta \, d\theta. \tag{4}$$

Since the term $\sin\theta$ in Eq. (4) corresponds to the theoretical probability density function of the sound incident angle distribution in the perfectly diffuse sound field, the ''effective''statistical absorption coefficient in a room can be defined by modifying Eq. (4) as

TABLE III. Key as in Table II but α_p, α_s, and α_f are the absorption coefficients associated with S_p, S_s, and S_f areas, respectively.

		Frequency (Hz)					
		125	250	500	1k	2k	4k
Mitaka Concert Hall	α_p	0.33	0.44	0.81	0.93	1.06	1.12
	α_s	0.04	0.11	0.30	0.22	0.18	0.11
	α_f	0.12	0.20	0.27	0.17	0.13	0.07
Hamarikyu Asahi Hall	α_p	0.46	0.59	0.87	0.98	1.12	1.08
	α_s	0.16	0.30	0.32	0.21	0.11	0.12
	α_f	0.09	0.20	0.28	0.24	0.10	0.05
New National Theater Medium Theater	α_p	0.37	0.55	0.80	0.89	0.96	1.00
	α_s	0.12	0.10	0.36	0.30	0.24	0.21
	α_f	0.18	0.24	0.29	0.19	0.16	0.15
New National Theater Opera House	α_p	0.42	0.54	0.77	0.89	0.99	1.07
	α_s	0.00	0.03	0.38	0.33	0.25	0.23
	α_f	0.17	0.14	0.31	0.26	0.16	0.11
Vienna, Musikvereinssaal	α_p	0.30	0.38	0.79	0.97	1.13	1.12
	α_s	0.00	0.10	0.24	0.21	0.13	0.13
	α_f	0.05	0.23	0.28	0.16	0.06	0.11
Tokyo Opera City Concert Hall	α_p	0.37	0.43	0.75	0.87	0.99	1.06
	α_s	0.01	0.04	0.31	0.30	0.18	0.20
	α_f	0.08	0.15	0.26	0.25	0.09	0.14

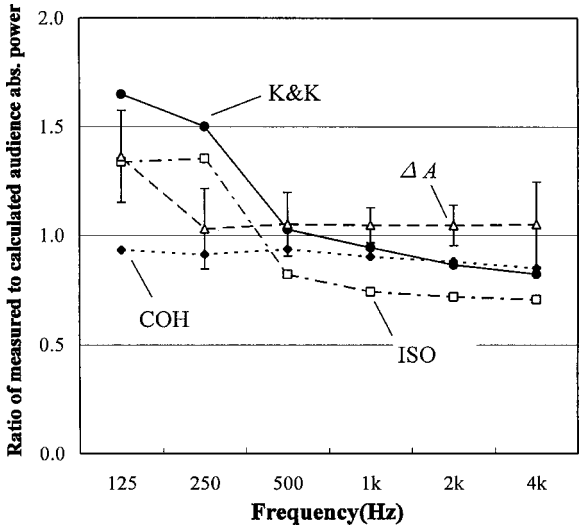


FIG. 3. Average values of the ratios of measured to calculated audience absorption power. The former is derived from measured reverberation times and the latter is predicted by four methods, ISO, K&K, COH (Beranek, 1996; Beranek and Hidaka, 1998) and ΔA (total absorption per person), for the six halls shown in Table II. The vertical bars indicate the variation range of the ratio by ΔA method, for which the spread is substantially more than that for the other three methods.

$$\alpha_{\text{field}}=\frac{\int_0^{\pi/2}\alpha(\theta)\cos\theta P(\theta)\,d\theta}{\int_0^{\pi/2}\cos\theta P(\theta)\,d\theta}, \tag{5}$$

where $P(\theta)$ is the probability density function for the actual hall and θ means the incident angle of each sound wave arriving at a specified receiving point.

The shape of the reverberation chamber in Takenaka R&D was precisely reconstructed in the computer as a CAD model, including the 20 diffusion panels which were randomly suspended in the space. An audience area equivalent to $2.8\times3.8\,\text{m}^2$ was marked off on the floor. A "receiver" was located in the center of this area. In the calculation, the audience area was assumed to be perfectly absorptive and the height of the receiving point was 1.2 m. The function $P(\theta)$ was calculated using the ray tracing method, where the number of rays traced was 100 000 and the maximum reflection order was up to 40 or the maximum time up to 1000 ms.

Figure 4 shows the numerical results for the Takenaka reverberation chamber, where the effective sound energy distributions per unit area, $\cos\theta P(\theta)$, in both the vertical and horizontal planes are plotted against the incident angle θ. The solid line gives $\cos\theta P(\theta)$ for every degree and the solid points are for $\cos\theta\{\Sigma P(\theta)\}$, calculated every 15 degrees. In the vertical plane, it is seen that the values for the CAD model behave similar to the theoretical value for perfect diffusion ($\sin 2\theta$, dotted line).

These same measurements were made in a CAD model of Boston Symphony Hall ($V=18\,750\,\text{m}^3$), a typical shoebox-shaped hall (see Fig. 5). Perfect sound absorption for the audience areas was assumed at all frequencies. By comparison with Fig. 4, $\cos\theta P(\theta)$ vs. θ reveals an uneven distribution, particularly approaching grazing incidence (above 50 degrees) in the vertical plane and at lateral incidence (particularly ±45 to ±75 degrees) in the horizontal plane. In this experiment, the receiving position was near the

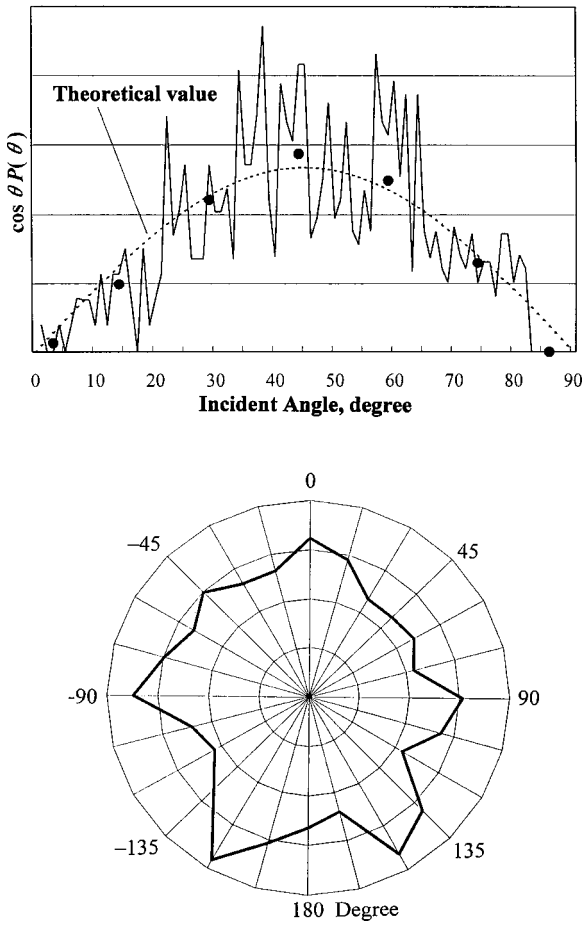
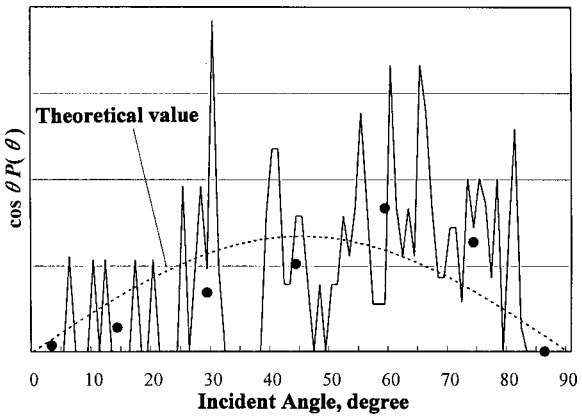


FIG. 4. Typical example of the sound energy distribution per degree and per unit area as a function of (upper) the incident angle in the median plane, and (lower) averaged distribution at intervals of 15 degrees in horizontal plane. The data were obtained in a CAD model of the reverberation chamber of Takenaka R&D. A sound absorbing layer ($2.8\times3.8\,\text{m}^2$) was placed at the center of the floor of the "chamber." The dotted line is the theoretical value, $\sin 2\theta$, that assumes perfect diffusion and the black dots mean the average distribution at intervals of 15 degrees in the vertical plane. The sound source ($h=0.5$ m) is placed at the corner of the chamber and the receiving point ($h=1.2$ m) is in the center of sound absorbing object.

fore-aft center of the main floor, but was shifted to the side of the center line by a few meters. This difference was found to be more apparent under a balcony or at seats on the upper balcony, indicating an abundance of deviated reflections from specific directions, as one would expect. This result shows that the (effective) statistical absorption coefficients in actual occupied halls are different from those in a reverberation chamber with ideal diffusion.

As a supplementary discussion, these results may depend on the applicability of the ray tracing method in the frequency range of 125 to 250 Hz. The validity of the results must be associated with the existence of a sufficient number of modes in each frequency range as long as each surface area is properly subdivided in relation to the wavelength. In a room, the average density of eigenfrequencies is approximately equal to $4\pi Vf^2/c^3$, where V and c are room volume and sound velocity, respectively. Taking into account the values of V above, i.e., 18 750 m^3 for the hall and 332.8 m^3 for the reverberation, the number of eigenfrequencies at 63 Hz in the hall is 23, which is similar in magnitude to that at 500 Hz



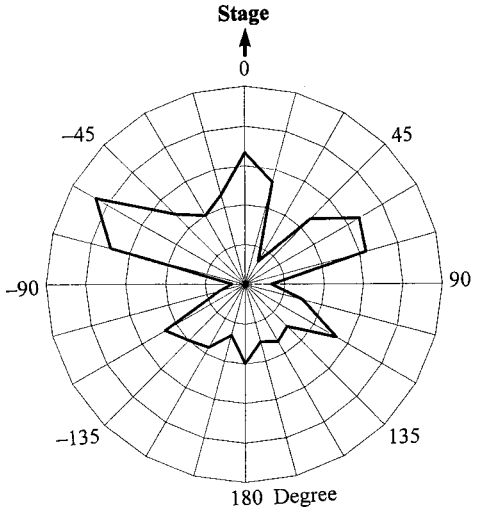


FIG. 5. Key as in Fig. 4, but for the model of Boston Symphony Hall. The sound source position is 3 m from the stage edge on the center line at a height of 1.5 m, and the receiving position is near the center of the main floor but shifted to the side a few meters.

in the reverberation chamber, i.e., 26. This implies that the sound field in the hall at 63 Hz is equivalent to that in the reverberation chamber at 500 Hz. Next, note the ''Schroeder frequency'' (Schroeder, 1954) that defines the lower limiting frequency at which a statistical treatment is permissible. With it, letting T equal the reverberation time, $f_s \cong 2000\sqrt{T/V}$. For the hall ($T=2$ s), $f_s=21$ Hz, and for the reverberation chamber ($T=12$ s), $f_s=380$ Hz.

IV. OBLIQUE ABSORPTION COEFFICIENT OF AN AUDIENCE

Even if $P(\theta)$ for a particular hall were known, another variable that is needed to evaluate Eq. (5) is the oblique absorption coefficient of the audience, $\alpha(\theta)$. If the preponderance of distribution of sound rays in a hall approaches grazing incidence (see Fig. 5, upper, above 50 degrees), α_{field} in a hall will certainly be different from α_{stat} in a reverberation chamber.

Figure 6 is a numerical example of $\alpha(\theta)$ for a homogeneous plane porous material, 25 mm in thickness, with rigid backing [see also Fig. 6 in Hunt (1939)]. The absorption coefficient is greater than 0.6 at frequencies of 1 kHz and above and the variation is gradual as a function of the incident angle, at least up to an angle of 70 degrees. At frequencies below 500 Hz the absorption coefficient is smallest at normal incidence and increases rapidly as the angle gets closer to 80 degrees. We believe that one can expect a similar tendency for the audience absorption. If so, it follows that α_{field} at mid- and high frequencies is less affected by the degree of sound diffusion in the sound field, because there $\alpha(\theta)$ is less dependent on the incident angle. At low frequencies the value of α_{field} are highly dependent on the incident angle distribution for individual rooms. This hypothesis is consistent with the measurements of Fig. 3 which show that the greatest differences between the reverberation-chamber absorptive powers of occupied chairs by the K & K and ISO methods and those measured in actual halls are significantly greater at low frequencies.

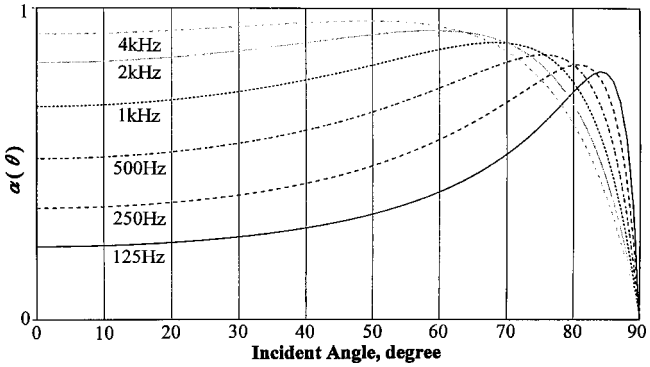


FIG. 6. Example of the oblique sound absorption coefficients for a homogeneous plane porous layer with rigid backing at 1/1 octave center frequency of 125 to 4 kHz (flow resistance per cm: 65 $g\cdot s^{-1}\cdot cm^{-3}$; thickness: 25 mm).

A. New measuring method of oblique sound absorption coefficients

Oblique sound absorption coefficients are difficult to measure, especially for an audience, and no established procedure is available, although such absorption coefficients would appear to be one of the essential factors in room

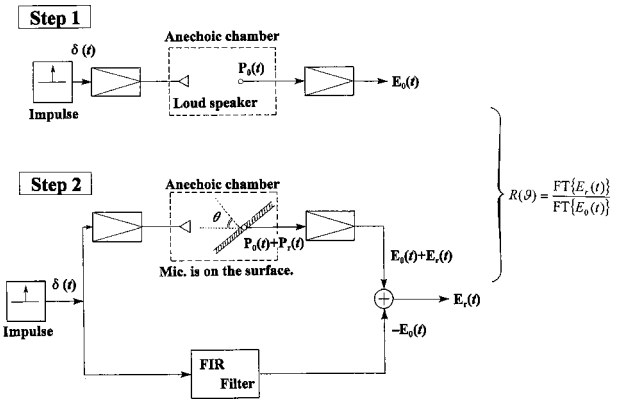


FIG. 7. Block diagram of the new measuring method of oblique absorption coefficient by digital signal processing. In step 1, the direct sound, $E_0(t)$, is measured in an anechoic chamber with no sample, which is used as the coefficient of a FIR filter. Next, the reflected signal, $E_r(t)$, is extracted from the observed signal, $E_0(t)+E_r(t)$, measured with samples by adding the electrically generated signal with the inverse phase, $-E_0(t)$, using the FIR filter in step 2. Note that the microphone must be at the surface of the sample and that the distance between the loudspeaker and the microphone must be the exactly same for both steps.

acoustics. Several methods of measurement have been reported (Hunt, 1939; Ingard and Bolt, 1951; Hirata, 1972), but the accuracy is known to deteriorate as the incident angle approaches that of grazing.

The seated audience in a hall for music is a collection of separated bodies a meter and a half in depth (floor to top of head). Because the audience absorption coefficients for sound fields approaching grazing incidence are very important, any existing method of obtaining them is inappropriate. We are reporting here a new laboratory measurement procedure, using digital processing, for determining the sound absorptions of an inhomogeneous-in-depth sample, such as a seated audience. The primary characteristic of this method is that for any sample it eliminates the direct sound and retains the reflected sound. In principle, this method is equal in accuracy for any incident angle, and it can be applied to objects with irregular surfaces or shapes if, for them, a hypothetical plane surface is defined over which the absorptions of the parts of the array can be collected. This definition of oblique absorption coefficient is not as exact as that for acoustic materials with flat surfaces; however, it should be noted that, in practice, an audience area is assumed to be a flat plane with a certain height. This assumption is the same for all four of the RT calculation procedures.

The procedure is shown in Fig. 7. As the first step, the impulsive direct sound, $E_0(t)$, generated by a loudspeaker is measured in an anechoic chamber by a microphone at distance L (with no object present) (see step 1). The waveform, $E_0(t)$, is stored as a coefficient of the finite impulse response, FIR, filter in step 2. In the upper part of step 2, the same sound signal is generated with a sample installed. The microphone must be on the surface of the sample and the distance L between it and the loudspeaker must be exactly the same as it was in step 1 for any angle. The observed waveform is $E_0(t)+E_r(t)$, where the latter means reflected waveform. Finally, the electrical signal $-E_0(t)$ with a delay time corresponding to the spacing, L, between the loudspeaker and microphone is generated from the FIR filter, and both signals are summed to yield the reflected term, $E_r(t)$. In this experiment, the pressure reflection coefficient, $R(\theta)$, as a function of frequency, is given by the following equation, if the microphone is placed at the surface of the material:

$$R(\theta)=\frac{\mathrm{FT}\{E_r(t)\}}{\mathrm{FT}\{E_0(t)\}}, \tag{6}$$

where FT represents Fourier transformation. In our experiment, L was 2 m and the sampling frequency was 22.05 kHz. The application of this measuring method for any porous material is presented in the Appendix.

B. Measurement of oblique absorption coefficient of audience

This new type of measurement was executed for a seated audience in the anechoic chamber of Takenaka R&D [8.4 m(l)×7.8 m(w)×9.8 m(h)]. Eighteen chairs (6 chairs ×3 rows, distance between each row: 950 mm) held mannequins, and were supported by a wooden board (12 mm thickness) laid on the otherwise acoustically transparent floor

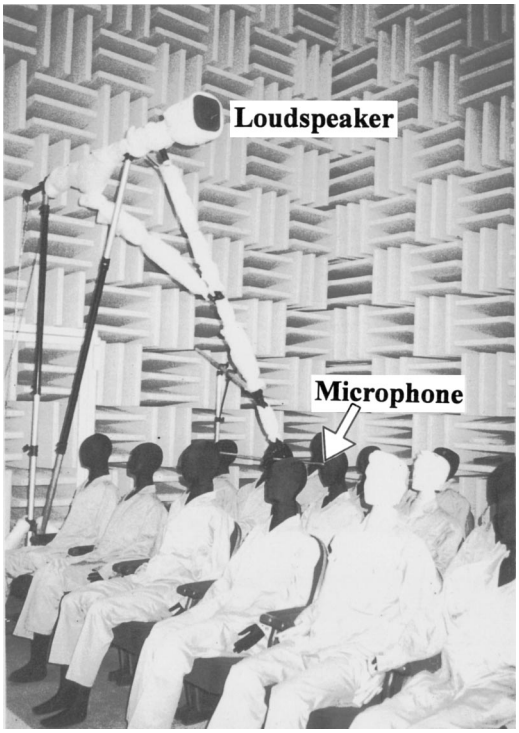


FIG. 8. Photograph of the configuration for the measurement of the oblique absorption coefficients for an audience seated in chairs on a plywood board in the anechoic chamber of Takenaka R&D. Eighteen mannequins dressed in proper clothes simulated a live audience. Note that the microphone can be placed on any part of a mannequin's body. The loudspeaker is rigidly held at 2-m distance from the microphone by a pole (shown covered with sound-absorbing material).

(Fig. 8). The total sound absorption of each mannequin with normal clothes was chosen to equal that of an average person, α_s, as determined by prior measurements in the reverberation chamber.

Each mannequin and seat was subdivided into nine parts (head, face, shoulder, breast, abdomen, thigh, shank, feet, and seat back shown by 15 numbers) and the remaining areas, that is, the bare floor and the uncovered seatback (shaded) (see Fig. 9). Fifteen measuring points plus the two shaded areas were defined as the representative ones for $\alpha_i(\theta_j)$ $(i=1,17)$. The assumption here is that this simulated seated audience is equivalent to a rectangular-parallelepiped block just as had been assumed in our earlier determinations of RTs in a hall or a reverberation chamber. The total absorption coefficient for the top surface of this block, as shown by

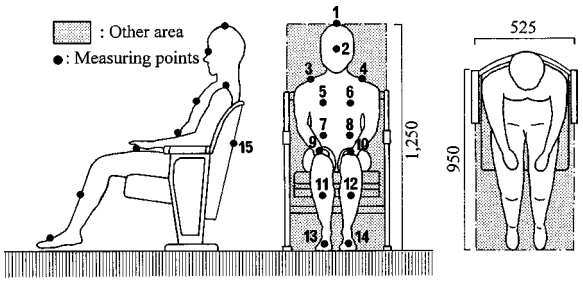


FIG. 9. Occupied area per seated person and the definition of the representative measuring points (●) in each part, i.e., head, face, shoulder, breast, abdomen, thigh, shank, feet, and seat back, and the other two areas shown by the hatching.

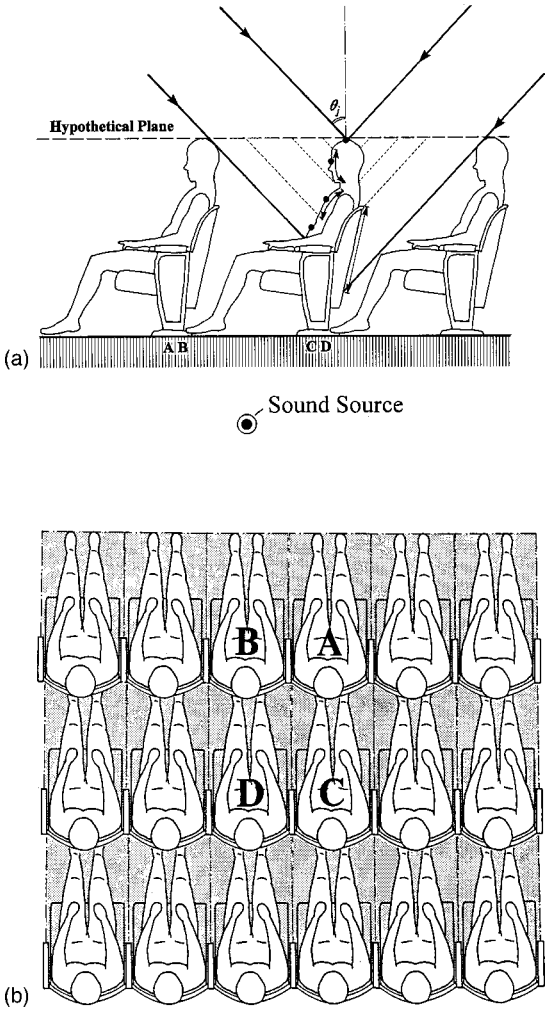


FIG. 10. (a) Definition of the incident angle θ_j. The hypothetical plane, at which the oblique incident absorption coefficient is defined, is at the head height of the seated audience. The black dots mean receiving points where the microphone was located. The equivalent surface area of face, breast, and seat back are shown as an example. A, B, C, and D mean the measuring seats that were placed at the front and second rows of a total of 6 chairs $\times 3$ rows. (b) The seated mannequins in the 18 chairs. Seats A–B and C–D were measured separately to observe the influence of seats in front of others.

the hypothetical plane in Fig. 10(a), is given by

$$\alpha(\theta_j) = \frac{\Sigma_{i=1}^{n} \alpha_i(\theta_j) \cdot S_i}{\Sigma_{i=1}^{n} S_i}, \tag{7}$$

where S_i is the area of each part of the seated audience reached by the incident sound wave at an angle θ_j, and α_i is the absorption coefficient for that area. In the specific calculation of each area S_i, the human body was subdivided and replaced by several rounded shapes. Thus, S_i is the area of that surface of any one shape that, approximately, faces the direction of the incident wave. The areas so determined are shown in Table IV. When the sound wave impinges on the bare floor or uncovered (shaded) seatback directly, the corresponding projected surface S_i was defined as that perpendicular to the wave direction. We note that the microphone is placed successively at the 17 points, the distance between the source and the microphone is always L, and the angle θ_j is that between the vertical line and that connecting the sound source and the microphone at the measuring point on the mannequin.

To consider the influence of the front chairs, the measurement was executed for two mannequins (designated A and B) in the front row and two (C and D) at the second row [Fig. 10(b)]. Two types of chairs, one lightly upholstered (type used in Vienna's Musikvereinssalle, Wisner-Hager KG) and the other medium upholstered (type used in Vienna's Burgtheater, Wisner-Hager KG), were selected, which are typical for concert halls and opera houses (Fig. 11). The measurements were carried out at intervals of 15 degrees between 0 and 90 degrees and 0 and -90 degrees.

As an example, Fig. 12 compares $\alpha_i(45°)$ of the Burgtheater model at each point on the mannequins. It is seen that the oblique incident absorption coefficients at seats A and B have nearly the same values as those at seats C and D, except for the abdomen position. Next, Fig. 13 shows $\alpha_i(\theta_j)$ measured at the mannequin head as a function of θ_j. The absorption coefficient at 250 Hz is higher than that at 125 Hz and at 125 Hz it increases as the incident angle approaches grazing like that for a plane porous material (as Fig. 6). It is inter-

TABLE IV. Surface area S_i in ($\times 10^2$) cm^2 of each part of one mannequin in a seated audience corresponding to the incident angle θ_j onto the vertical plane defined in Fig. 10(a). The "other" areas are determined for any angle θ_j by projecting the (perpendicular and horizontal) shaded areas in Fig. 9 onto a surface at the front of the mannequin perpendicular to the direction of travel of the incident wave. The shaded areas are those that receive energy from the incident wave that is not blocked by a mannequin. The numbers in parentheses are the percentage ratios of S_i's to total areas.

θ_j (deg)	Head	Face	Shoulder	Breast	Abdomen	Thigh	Shank	Feet	Seat back	Other	**Sum**
90	4.3									2.6	**6.9**
	(63)									(37)	(100)
75	4.3	7.9								8.6	**20.8**
	(21)	(38)								(41)	(100)
60	4.3	7.9	6.1							10.9	**29.2**
	(15)	(27)	(21)							(37)	(100)
45	4.3	7.9	6.1	19.3	9.0				14.9	12.6	**74.1**
	(6)	(11)	(8)	(26)	(12)				(20)	(17)	(100)
30	4.3	7.9	6.1	19.3	9.0	14.4			17.6	12.8	**91.4**
	(5)	(9)	(7)	(21)	(10)	(16)			(19)	(14)	(100)
15	4.3	7.9	6.1	19.3	9.0	14.4	10.4		37.4	20.8	**129.6**
	(3)	(6)	(5)	(15)	(7)	(11)	(8)		(29)	(16)	(100)
0	4.3	7.9	6.1	19.3	9.0	14.4	10.4	3.9	47.3	20.5	**143.1**
	(3)	(6)	(4)	(13)	(6)	(10)	(7)	(3)	(33)	(14)	(100)

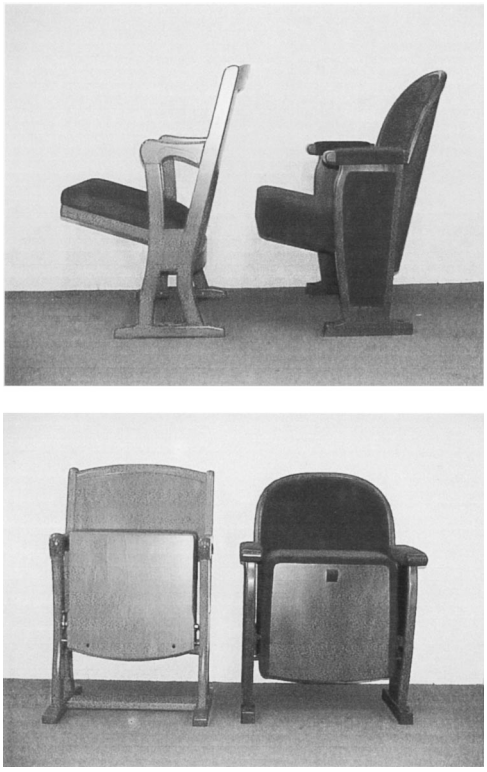


FIG. 11. Photographs of two types of audience chairs: Musikvereinssalle model (left in both photos) and Burgtheater model (right in both photos).

esting to see that the sound absorption of all the mannequin surfaces transferred to the hypothetical plane of Fig. 10(a) is, at 125 Hz, approximately the same (0.4 average below 60 degrees) as that of the porous material (see Fig. 6) with rigid backing (also 0.4 average below 60 degrees). Figure 14 is an example of a similar measurement of $\alpha_i(\theta_j)$ in the horizontal plane. The variation in this plane is not as extreme as that in

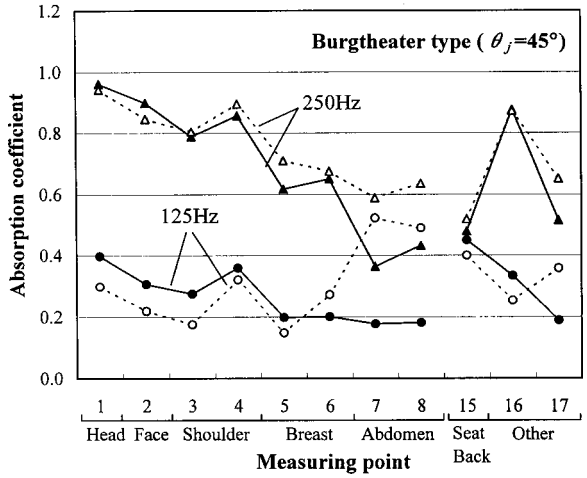


FIG. 12. Comparison of $\alpha_i(\theta_j)$ (defined at the hypothetical plane) for an audience in Burgtheater model chairs for $\theta_j=45$ degrees versus the measuring positions on the seated audience. The numbers on the x axis are those shown in Fig. 9. Eighteen mannequins (6 chairs×3 rows) formed the audience block. The measurement was executed for four mannequins, i.e., A and B, and C and D, which were placed, respectively, at the front (solid line) and the second row (dotted line). The averaged values for the two side-by-side seats at each row are shown in the figure. The ''thigh,'' ''shank,'' and ''feet'' absorptions are not shown because the sound wave does not strike there.

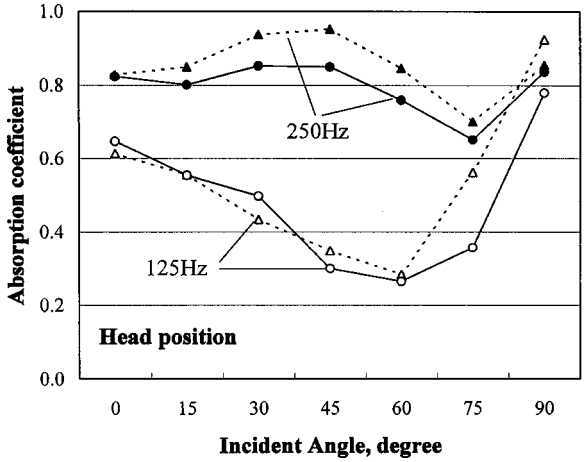


FIG. 13. Averaged $\alpha_i(\theta_j)$ at the head position of four mannequins at the front and second rows for Musikvereinssalle (solid line) and Burgtheater (dotted line) model chairs as a function of incident angle θ_j in the median plane.

the vertical plane. Obviously, a three-dimensional measurement of $\alpha_i(\theta_j)$ is a time-consuming procedure, but, from comparison of Figs. 12 and 14, one can assume that $\alpha_i(\theta_j)$ determined by moving the loudspeaker around in a circle on a horizontal plane (angle θ_j held constant) is nearly the same at all points on the circle. In Fig. 15 it is seen that the values of $\alpha(\theta_j)$ [see Eq. (7)], the absorption coefficients of the hypothetical plane, increase as the incident angle θ_j approaches grazing incidence, and this characteristic resembles that of the behavior of a porous material (Fig. 6). This finding confirms our assumption in the previous section, which was based on Fig. 6.

V. ABSORPTION COEFFICIENT OF AN AUDIENCE IN A REVERBERATION CHAMBER AND IN CAD MODELS OF HALLS FOR MUSIC

The effective statistical absorption coefficient of Eq. (5) that reflects the practical sound incidence distribution in the sound field is rewritten in the discrete form

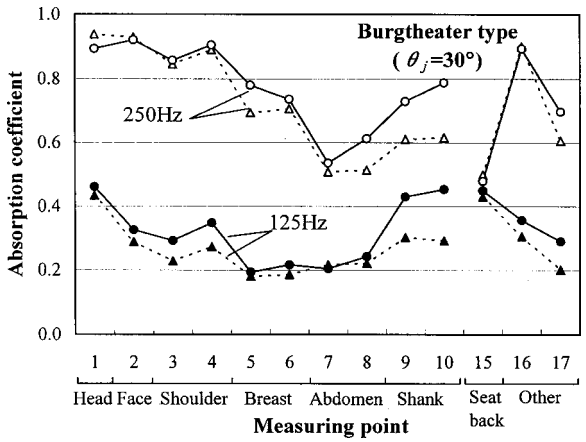


FIG. 14. Comparison of $\alpha_i(30°)$ of Burgtheater model chair in the vertical plane, but for 45 degrees (solid line) and 0 degrees (dotted line) in horizontal plane. The numbers along the x axis are those shown in Fig. 9. The ''shank'' and ''feet'' absorptions are not shown because the sound wave does not strike there.

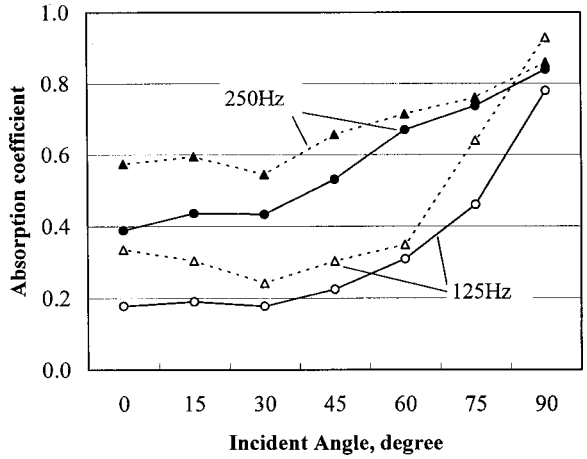


FIG. 15. Oblique absorption coefficients at the hypothetical plane of a seated audience, at the head height for Musikvereinssalle (solid line) and Burgtheater (dotted line) model chairs, based on measurements of Fig. 8.

$$\alpha_{\text{field}}=\frac{\Sigma_{\theta_j=0}^{\pi/2}\alpha(\theta_j)\cos\theta_j P(\theta_j)}{\Sigma_{\theta_j=0}^{\pi/2}\cos\theta_j P(\theta_j)}. \tag{8}$$

The probability density function $P(\theta_j)$ in this equation was determined by the CAD model of the reverberation chamber described in Sec. III. As a precaution, α_{field} was obtained for five receiving points in the reverberation chamber, and the numerical differences among them were very small.

Table V is a comparison of α_s, α_{stat}, and α_{field} determined in the reverberation chamber for seated audiences on the two kinds of the chairs of Fig. 11. Here, α_s was measured using 20 chairs (5 chairs×4 rows) with stiff barriers (height: 1.2 m) around them. The three absorption coefficients for each chair are in good agreement with each other as the difference between α_s and α_{field} is within acceptable ranges. In this case, α_{field} is in good agreement with α_{stat} since the diffusivity of the sound field in the Takenaka reverberation chamber is adequately large even at these low frequencies. This result indicates that both equations, namely, (7) and (8), yield the same measured values α_s in the reverberation chamber.

The same comparison of (7) and (8) was carried out for five concert halls (see Table VI) using CAD models: Tokyo, Hamarikyu-Asahi, Boston Symphony Hall, Vienna Musikvereinssaal, Tokyo Opera City Concert Hall, and Mitaka Concert Hall. Their architectural/acoustical features are described in Beranek (1996) and Hidaka *et al.* (2000, 2001).

TABLE V. Comparison of the three kinds of absorption coefficients, α_s, α_{stat}, and α_{field}. The coefficient α_s was measured in the Takenaka reverberation chamber by the ISO method (with stiff barriers around seated mannequins). α_{stat} and α_{field} were calculated by Eqs. (4) and (8) where $\alpha(\theta_j)$ was measured in the anechoic chamber (see Fig. 8) and $P(\theta_j)$ was measured in the CAD model (see Sec. III).

	Musikverein type		Burgtheater type	
Hz	125	250	125	250
α_s	0.26	0.44	0.43	0.64
α_{stat}	0.26	0.56	0.35	0.65
α_{field}	0.27	0.56	0.35	0.66

TABLE VI. Five concert halls used for the calculations in this study. V is room volume; S is the total room surface area including that of the audience area.

	V (m^3)	S (m^2)	Mean free path (m)	No. of Surfaces
Mitaka Concert Hall	5500	2316	9.5	657
Hamarikyu Asahi Hall	5800	2570	9.0	1990
Vienna Musikvereinssaal	15 000	4102	14.6	383
TOC Concert Hall	15 300	5843	10.5	1187
Boston Symphony Hall	18 750	5778	13.0	861

The values of $\cos\theta_j P(\theta_j)$ were obtained, as for Figs. 4 and 5. The computer CAD models were constructed precisely from architectural drawings. The surface details of the models were sized to fit the wavelengths at 125 and 250 Hz. While the number of surfaces needed to constitute a CAD model varies from hall to hall, the conditions that were used to calculate $P(\theta_j)$ for all five halls are (1) about 0.1 to 0.8 surface per m^2 (total number of surfaces divided by the total surface area), (2) perfect absorption at the audience area, (3) height of receiver is 1.2 m from the floor, and (4) sound source position is 3 m behind the stage edge on the center line and its height is 1.5 m. In the analysis of α_{field} [see Eq. (8)], more than 12 receiving positions were chosen for the calculation and averaged. The standard deviation of α_{field} was about 0.02 for 125 and 250 Hz.

As the next step, the edge effect must be taken into account if one is to compare α_{field} and α_{hall}. Here, α_{hall} means the absorption coefficient of a seated audience calculated from RT values measured in the five halls and α_{field} are the calculated values using the measured $P(\theta_j)$ quantities for each hall.

The mean squared sound pressure $\langle p_r^2\rangle$ at a point located at a vertical distance x from a rigid surface in a perfectly diffuse sound field of a reverberation chamber (Waterhouse, 1955) is

$$\langle p_r^2\rangle \propto \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\pi/2}(1+\cos(2kx\cdot\cos\theta))\cdot\sin\theta\,d\theta\,d\varphi$$
$$=1+\frac{\sin(2kx)}{2kx}, \tag{9}$$

where $k=2\pi/\lambda$. For $x\gg\lambda$, the incremental value of squared sound pressure from 0 to x is

$$\Delta p^2\propto\int_0^x(\langle p_r^2\rangle-1)\,dx\approx\int_0^\infty(\langle p_r^2\rangle-1)\,dx=\frac{\lambda}{8}. \tag{10}$$

Because perfect diffusion is almost fully established in the reverberation chamber, the sound pressure increase is calculated as $\lambda/8$ times the surface area of the chamber if the effects of edge and corner of the wall surfaces of Fig. 2 are ignored (Kath and Kuhl, 1961).

On the other hand, the sound incidence distribution differs in a hall for music as discussed before. Equations (9) and (10) must be modified by replacing the $\sin\theta$ term with $P(\theta)$. Then, one obtains

TABLE VII. Comparison of measured and calculated sound absorption coefficients for five halls, α_{hall} and α_{field}, which were obtained, respectively, from (a) measured values of RT under occupied conditions and (b) the effective absorption coefficient by Eq. (8). In the former, the corrections by modified Waterhouse edge effect, that is, λ/3 addition to the audience area, are included.

	Hz	α_{hall}	α_{field}
Hamarikyu Asahi Hall	125	0.48	0.41
	250	0.73	0.69
Boston Symphony Hall	125	0.52	0.31
	250	0.71	0.61
Vienna Musikvereinssaal	125	0.47	0.38
	250	0.64	0.68
TOC Concert Hall	125	0.41	0.31
	250	0.61	0.61
Mitaka Concert Hall	125	0.48	0.32
	250	0.63	0.63

$$\Delta p^2 \propto \int_0^x (\langle p_r^2 \rangle - 1)\, dx$$

$$= \int_0^x \int_0^{\pi/2} \cos(2kx \cdot \cos\theta) \cdot P(\theta)\, d\theta\, dx, \tag{11}$$

and, by rewriting in the discrete form for numerical purpose,

$$\Delta p^2 \propto \sum_{\theta_j=0}^{\pi/2} P(\theta_j) \sum_{x_m=0}^{x} \cos(2kx_m \cdot \cos\theta_j) \cdot \Delta x \cdot \Delta\theta,$$

$$\Delta x = 0.5\lambda, \quad \Delta\theta = 15 \text{ degrees}. \tag{12}$$

The sound pressure increase Δp^2 was evaluated in the CAD model by using Eq. (12) substituting $P(\theta_j)$ for each hall (see Fig. 5 for one hall), and, as a result, it was found that the average sound pressure increase in the five concert halls listed in Table VI rises about $\frac{13}{5}$ times that in the perfectly diffuse sound field, that is, about λ/3 instead of λ/8 of Eq. (10) for the reverberation chamber. Now, these values for 125 and 250 Hz are 90 and 50 cm, respectively. If converted to the audience surface area in order to calculate the absorption coefficient of the audience in the hall, they correspond to the total floor area (using full area of aisles and other areas) at 125 Hz for many cases and the area defined by the COH method at 250 Hz. In other words, the empirical definition of audience areas by the edge effect by Beranek (1960) (0.5 m) has been proved at 250 Hz theoretically.

By using this definition of the audience area, the absorption coefficient of the audience in each hall was calculated and compared with actual value α_{hall} in Table VII. The results show that they are in satisfactorily agreement with each other, both at 125 and 250 Hz excluding that at 125 Hz in Boston Symphony Hall.[2] That is to say, the discrepancy of the K & K method at low frequencies as shown by Fig. 3 was solved by the new measuring method and definition of audience area, simply by the increase in the audience area S_p due to the added 90-cm edge strip at 125 Hz and 50-cm edge strip at 250 Hz. Figure 16 shows the average values of the ratios of the measured to the calculated total absorption power of the seated audiences by K & K method and the new method in the five halls shown in Table VI. Accordingly, one can obtain the absorption coefficient of an audience by using the K & K method as accurately as the COH method in Fig. 3. Furthermore, this new method is applicable for the prediction of the RT of a particular hall, where chairs with a new type of design are contemplated, and the COH method is not applicable.

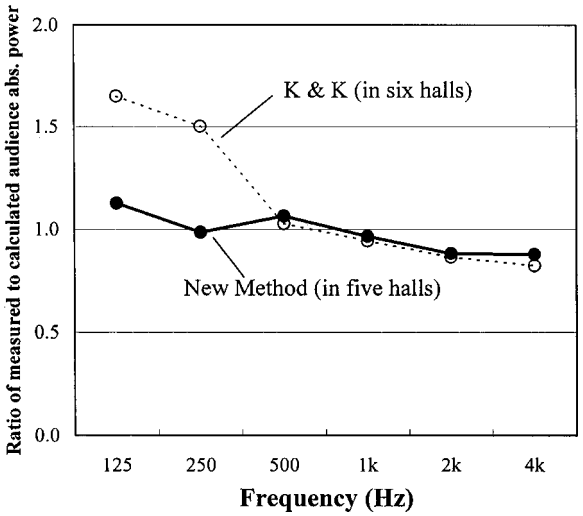


FIG. 16. Comparison of the average of the ratios of audience absorption power derived from measured reverberation times to those predicted by K & K method in the six halls shown in Table II and the new method in the five halls shown in Table VI at low frequencies.

VI. EXAMPLE

The Tokyo Opera City (TOC) Concert Hall of Tables II, III, and VI is used as an example. The sequence of calculations and the numerical results are shown in Table VIII. From 500 to 4000 Hz, the K & K reverberation-chamber method was employed for TOC, as discussed in Sec. II A 2, because the oblique absorption coefficients of a seated audience in a large room are less dependent on the incident angles of the traveling waves than at lower frequencies. For the two lower frequencies the procedure differs from the K & K method. The logical progression is sketched briefly below. The lettering corresponds to that for the lines in Table VIII. In calculations of reverberation times at frequencies above 500 Hz, the sound-absorbing power of the air must also be included.

(a) The sound absorption coefficients for the top, side, and front of the audience block, e.g., α_p, α_s, and α_f, as shown in Table III, were obtained in the reverberation chamber according to the K & K method. The particular seats chosen for TOC were used.

(b) From architectural plan-drawings of TOC, the areas S_p, S_s, and S_f were determined, as shown in Table II, and the products of the absorption coefficients in (a) and these areas were calculated and given for each of the frequencies.

(c) The actual sound absorbing power at each frequency as measured in the completed TOC is listed here. At the higher frequencies, the air absorption was subtracted out, before the numbers shown here.

(d) It is seen from the ratios of (c) to (b) presented in this

TABLE VIII. Example (Tokyo Opera City Concert Hall, TOC) of the procedure for reverberation chamber measurements of sound absorption coefficients by the K & K method above 500 Hz and new measurements at low frequencies to the prediction of audience absorption in halls for music.

		Frequency (Hz)					
		125	250	500	1k	2k	4k
(a) The absorption coefficients of top, side, and front area of audience block measured in the reverberation chamber by K & K method (from Table III).	α_p	0.37	0.43	0.75	0.87	0.99	1.06
	α_s	0.01	0.04	0.31	0.30	0.18	0.20
	α_f	0.08	0.15	0.26	0.25	0.09	0.14
(b) Total audience absorption, product of (a) and each area, $S_p=793$, $S_s=261$, and $S_f=39$, m^2 (from Table II).		299	357	686	778	836	898
(c) Total audience absorption measured in TOC using Eq. (1), m^2		466	614	650	705	711	744
(d) Ratio of (c) to (b)		1.56	1.72	0.95	0.91	0.85	0.83
(e) Calculation of absorption coefficient for TOC using Eq. (8)		0.31	0.61	···	···	···	···
(f) Areas for TOC used according to Sec. V (from Table II), m^2		1312	1052	···	···	···	···
(g) Product of (e) and (f), m^2		407	642	686	778	836	898
(h) Ratio of (c) to (g)		1.14	0.96	0.95	0.91	0.85	0.83

line that the K & K method underestimates the audience absorption in the 125- and 250-Hz bands.

(e) From the sound absorption coefficients $\alpha(\theta_j)$ measured in an anechoic chamber (see Fig. 8) and the probability density functions $P(\theta_j)$ determined in a CAD model, the ''effective'' absorption coefficients α_{field} [see Eq. (8)] were determined at 125 and 250 Hz. Obviously, this is the expensive part of the new method, because the procedure of Fig. 8 is required, and an accurate CAD model must be achieved.

(f) From the results of Eq. (12), the ''effective'' areas at 125 and 250 Hz are calculated from the floor plan. This turns out to be the equivalent of using the entire horizontal area under and around the seating, S_{ta}, at 125 Hz. For 250 Hz, the equivalent area is that over which the audience sits, plus 0.5 m around the edges (except when adjacent to walls or balcony fronts).

(g) Finally the ''corrected,'' predicted, total-audience absorbing powers for these frequencies are presented here, along with the higher frequency values by the K & K method taken from (b).

(h) Shown is a set of the ratios for TOC, one of the five halls whose ratios were averaged to obtain the lower curve of Fig. 16.

It is possible to predict RT with higher accuracy for new types of chairs than that predicted by the COH method for ''average'' chairs.

VII. CONCLUSIONS

To predict the reverberations times of a hall for music in the occupied condition, one must know the sound absorption per unit area by the audience. The accuracy of four existing methods, ISO, K & K, COH, and ΔA, for determining the sound absorption coefficients of an audience were compared with measurements in six halls. The COH method, which predicts the absorption coefficients of audiences from measurements made in actual halls, was found to be the most appropriate at all frequencies. The ISO method was least successful, while the ΔA method was good on average, although it led to widely varying values of reverberation times among the halls because it does not take into account the types of upholstering or chairs. Both the ISO and the K & K methods overestimated the reverberation times at low frequencies by a factor of 1.5 or more. The K & K method, which uses a reverberation chamber for the measurements, was most accurate at mid-frequencies. This article extends the accuracy of the K & K method to the low frequencies, a desirable goal because this method can then be used to determine audience absorption coefficients of new types of chairs or upholstering in a reverberation chamber.

It was surmised that the difference between audience absorption coefficients at low frequencies in reverberation chambers and those measured in halls for music was caused by the differences in the diffusion of their respective sound fields. The first stage of the investigation was to determine the actual state of diffusion, first, in a large reverberation chamber (with the audience test block in the center) and, second, in six halls fully occupied. It was found that the audience areas in occupied halls are bombarded by a greater percentage of sound waves near grazing incidence than in reverberation chambers (Figs. 4 and 5). Numerically, the probability density of the incident sound waves was determined.

In the next stage the absorption coefficients of a seated audience at 125 and 250 Hz as a function of angle of incidence of a sound wave was determined in an anechoic room by a novel procedure (Figs. 8–15). The absorption coefficients at these low frequencies increase greatly as the angle of incidence increases from 60 to 90 degrees.

Using the K & K method in the reverberation chamber, the absorption coefficients of the seated audience were determined by the audience area, which were the sum of the top, sides, and correction areas shown in Fig. 2. Using these coefficients, the reverberation times were calculated for six occupied halls and compared with those measured.

The comparison is shown by the upper (dotted) curve in Fig. 16.

Finally, it was shown mathematically that using both (a) the probability density of the incident sound waves as a function of angle of incidence and (b) the measured sound absorption coefficients of an audience as a function of angle of incidence, the sound energy (pressure squared) impinging on an audience is larger in the halls (proportional to $\lambda/3$) as compared to that in the reverberation chamber (proportional to $\lambda/8$), that is to say an amount of about $\frac{13}{5}$ larger. This energy increase can be accounted for in the calculations of the reverberation times in actual halls, simply by increasing the area of the audience in the halls. At 125 and 250 Hz, the results show that absorption coefficients determined by the K & K method may be improved (lower curve of Fig. 16) by an increase in area equal to a strip around each of the audience blocks in a concert hall of 90 cm at 125 Hz and 50 cm at 250 Hz (excluding the edges adjacent to walls or balcony fronts). It is believed that these corrections will apply to every type of seat. They will differ a little from one hall to another, but the results are at least as precise as the Beranek (1996) and Beranek and Hidaka methods (1998), which were found to be the most accurate of those methods studied before this procedure evolved.

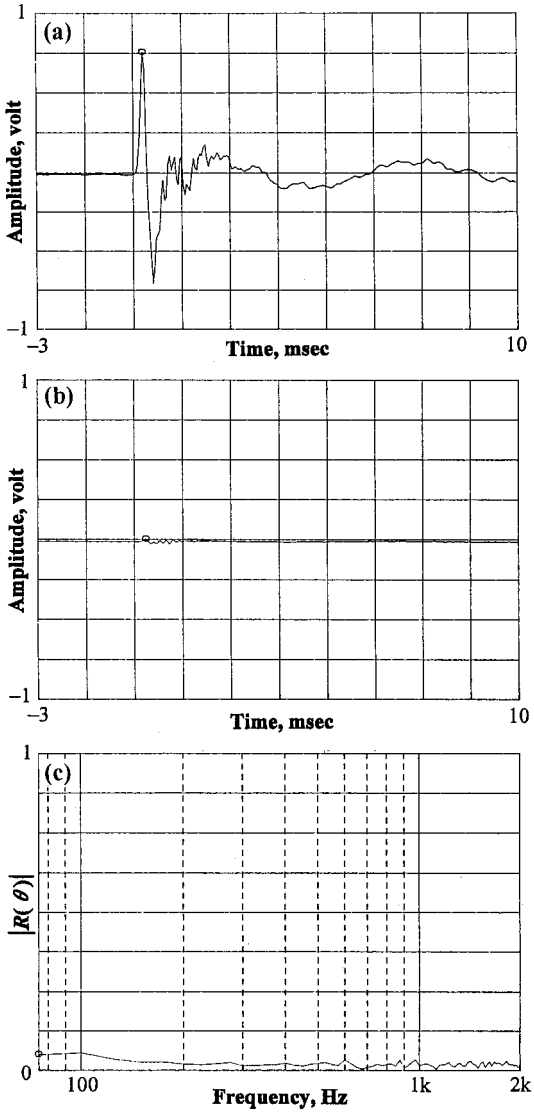


FIG. 17. Anechoic chamber test: (a) Waveform of direct sound $E_0(t)$ that is used for the coefficients of the finite impulse response (FIR) filter. (b) Residual value of the waveform by the operation $E_0'(t)-E_0(t)$, where $E_0'(t)$ is also the waveform measured without sample and $E_0(t)$ is that given by (a) above (c) $\mathrm{FT}\{E_0'(t)-E_0(t)\}/\mathrm{FT}\{E_0(t)\}$, which shows that the cancellation of the direct sound by this method function with enough accuracy for wide frequency range, because the residual error is less than 2% above 200 Hz.

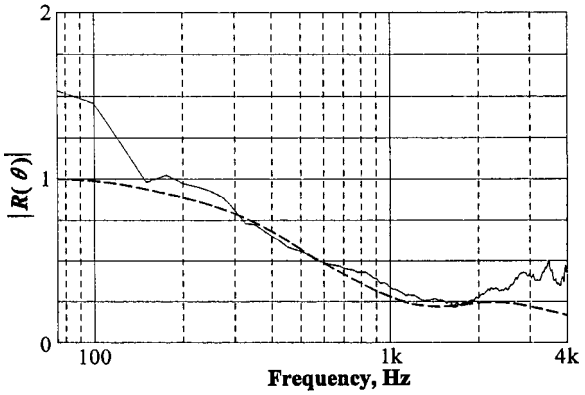


FIG. 18. Example of a measurement made, using this method, of the pressure reflection coefficient R for a glass wool sample 50 mm in thickness with rigid backing. The sound wave traveled perpendicular to the surface. The dashes are the calculated values based on the empirical equation by Delany and Bazley (1970) with flow resistance per cm, 24 $\mathrm{g\cdot s^{-1}\cdot cm^{-3}}$, which is the certified value of the test sample. The anomalous value (above one) at low frequencies is caused by edge diffraction, because the sample size was limited to 910×1820 mm. For measurement of audience chairs in this article, the sample size was large enough to avoid this phenomenon.

APPENDIX: AN APPLICATION EXAMPLE OF NEW MEASURING METHOD OF OBLIQUE ABSORPTION COEFFICIENT

As a preliminary experiment, the measurement was executed in an anechoic chamber so that the numerical error that will result from elimination of direct sound was evaluated. Figure 17 shows (1) the direct sound, $E_0(t)$, measured at step 1 in Fig. 7, (2) residual value of the waveform by the operation $E_0'(t)-E_0(t)$, where $E_0'(t)$ is the waveform measured with no sample at different air condition, and (3) $\mathrm{FT}\{E_0'(t)-E_0(t)\}/\mathrm{FT}\{E_0(t)\}$. It is impossible to completely eliminate direct sound signal since it involves effects of fluctuation of medium or digital rounded errors. Nevertheless, it is judged from the result of Fig. 17(c) that residual error by this cancellation method is less than 2% above 200 Hz, which is an acceptable value for usual measurement.

Figure 18 is a measurement example of the normal incident reflection coefficient for plane porous materials, i.e., glass wool of 50 mm in thickness, with rigid backing. The measurement demonstrated that this method could give stable results for all frequency ranges. It should be noted that perfect elimination of direct sound requires adequate control of temperature and humidity in an anechoic chamber, which was done in these tests.

[1]For each hall, two sets of seats were measured, i.e., the manufacturer's samples and the seats actually installed. The sample seats were measured to decide on the design, but a problem arose with the installed seats. In three of the six halls, the upholstery cloth on the installed seats was backsprayed. The coefficients reported here are for the seats as actually installed.

[2]For Boston Symphony Hall, the absorption of the elevated wooden floor was included as part of the seated audience absorption.

ASTM C423-90a (**1990**). ''Test method for sound absorption and sound absorption coefficients by the reverberation room method'' (American Society for Testing and Materials Standards).

Balachandran, C. G., and Robinson, D. W. (**1967,1968**). ''Diffusion of the Decaying Sound Field,'' Acustica **19**, 245–257.

Beranek, L. L. (**1960**). ''Audience and seat absorption in large halls,'' J. Acoust. Soc. Am. **32**, 661–670.

Beranek, L. L. (**1962**). *Music, Acoustics and Architecture* (Wiley, New York).

Beranek, L. L. (**1969**). ''Audience and seat absorption in large halls. II,'' J. Acoust. Soc. Am. **45**, 13–19.

Beranek, L. L. (**1996**). *Concert and Opera Halls: How They Sound* (Acoustical Society of America, Melville, NY).

Beranek, L. L., and Hidaka, T. (**1998**). ''Sound absorption concert halls by seats, occupied and unoccupied, and by the hall's interior surfaces,'' J. Acoust. Soc. Am. **104**, 3169–3177.

Beranek, L. L., Hidaka, T., and Masuda, S. (**2000**). ''Acoustical design of the opera house of the New National Theater, Tokyo, Japan,'' J. Acoust. Soc. Am. **107**, 355–367.

Bradley, J. S. (**1992**). ''Predicting theater chair absorption from reverberation room measurements,'' J. Acoust. Soc. Am. **91**, 1514–1524.

Bradley, J. S. (**1996**). ''The sound absorption of occupied auditorium seating'' J. Acoust. Soc. Am. **99**, 990–995.

BS 3638 (**1987**). ''Method for measurement of sound absorption in a reverberation room.''

Dämmig, P. (**1991**). ''Model Investigation into Sound Fields in Reverberation Rooms,'' Acustica **75**, 105–120.

Davis, W. J., Orlowski, R. J., and Lam, Y. W. (**1994**). ''Measuring auditorium seat absorption,'' J. Acoust. Soc. Am. **96**, 879–888.

Delany, M. E., and Bazley, E. N. (**1970**). ''Acoustical Properties of Fibrous Absorbent Materials,'' Appl. Acoust. **3**, 105–116.

DIN 52212 (**1961**). ''Testing of architectural acoustics; measurement of sound absorption coefficient in a reverberation room'' (German National Standards).

Furrer, W. (**1964**). *Room and Building Acoustics and Noise Abatement* (Butterworths, London), pp. 70–71.

Hidaka, T., Nishihara, N., and Beranek, L. L. (**1996**). ''Mechanism of sound absorption by seated audiences in concert halls,'' Presented at the 3rd Joint Meeting: Acoustical Societies of America and Japan, J. Acoust. Soc. Am. **100**, 2705(A).

Hidaka, T., Nishihara, N., and Beranek, L. L. (**2001**). ''Relation of acoustical parameters with and without audiences in concert halls and a simple method for simulating the occupied state,'' J. Acoust. Soc. Am. **109**, 1028–1042.

Hidaka, T., Beranek, L. L., Masuda, S., Nishihara, N., and Okano, T. (**2000**). ''Acoustical design of the Tokyo Opera City (TOC) concert hall, Japan,'' J. Acoust. Soc. Am. **107**, 340–354.

Hirata, Y. (**1972**). ''Measurement of Absorption coefficient by electrically canceling method,'' J. Acoust. Soc. Jpn. **28**, 416 (in Japanese).

Hunt, F. V. (**1939**). ''Investigation of room acoustics by steady state transmission measurement, I,'' J. Acoust. Soc. Am. **10**, 216.

Ingard, U., and Bolt, R. H. (**1951**). ''A free field method of measuring the absorption coefficient of acoustic materials,'' J. Acoust. Soc. Am. **23**, 509.

ISO 354 (**1985**). ''Acoustics—Measurement of sound absorption in a reverberation room'' (International Organization for Standardization).

JIS A 1409 (**1998**). ''Method for measurement of sound absorption coefficients in a reverberation room'' (Japanese Industrial Standards).

Kath, U., and Kuhl, W. (**1961**). ''Einfluss von Streufläche und Hallraumdimensionen auf den Gemessenen Schallabsorptionsgrad,'' Acustica **11**, 50–64.

Kath, U., and Kuhl, W. (**1964**). ''Messungen zur Schallabsorption von Personen auf Ungepolsterten Stuhlen,'' Acustica **14**, 50–55.

Kath, U., and Kuhl, W. (**1965**). ''Messungen zur Schallabsorption von Polsterst hlen mit und ohen Personen,'' Acustica **15**, 127–131.

Kimura, S. (**1986**). *Architectural Acoustics and Noise Control* (Shoukoku-Sya, Tokyo), p. 201 (in Japanese).

Lord, P., and Templeton, D. (**1996**). *Detailing for Acoustics* (E&FN SPON), p. 195.

Makita, Y., and Hidaka, T. (**1987**). ''Revision of the cos θ Law of Oblique Incident Sound Energy and Modification of the Fundamental Formulations in Geometrical Acoustics in Accordance with the Revised Law,'' Acustica **63**, 163–173.

Nagata, M. (**1991**). *Acoustic Design of Architecture* (Ohmu-Sya, Tokyo), p. 215 (in Japanese).

Schroeder, M. R. (**1954**). ''Die statistischen Parameter der Frequenzkurven von großen Räumen,'' Acustica **4**, 594.

Waterhouse, R. V. (**1955**). ''Interference patterns in reverberant sound fields,'' J. Acoust. Soc. Am. **27**, 247–258.

Instantaneous frequency and short term Fourier transforms: Application to piano sounds

L. Rossi[a] and G. Girolami
UMR CNRS 6134, Systèmes Dynamiques Energétiques et Mécaniques, Equipe Ondes et Acoustique, Université de Corse, Faculté des Sciences, BP 52, 20250 Corte, France

(Received 1 July 1998; revised 1 August 2001; accepted 13 August 2001)

For more than 30 years and until nowadays, development of a system reproducing the functioning of human hearing has remained an aim difficult to reach. Recent methods for identification of the fundamental frequency of musical sounds obtain good results using information about the temporal evolution of the amplitude and frequency of individual sound partials. Piano sounds and polyphonic sounds, which can have several partials that are closely spaced in the frequency domain, have not been extensively tested by these procedures. In this paper, the Instantaneous Frequency (IF), as defined by the Hilbert transform, is used to obtain the frequency variations of piano sounds partials. The result implies that, for these sounds, the IF may contain modulations resulting in the separation of an apparent single sinusoid signal into two or more sinusoidal components at various times in the analysis process, which makes it impossible to use the temporal evolution of the frequency of partials for the procedure of note identification. The separation phenomenon also appears when the short term Fourier transform is used and can induce the detection of short-lived parasitic spectral peaks that must be taken into account by any note identification procedure based on the use of spectral information. [DOI: 10.1121/1.1409372]

PACS numbers: 43.60.Gk, 43.75.Yy [JCB]

I. INTRODUCTION

Piano sounds are usually analyzed in the frequency domain, where a decomposition into sinusoidal partials naturally occurs. One important problem with this kind of procedure arises from the nonstationary nature of these signals, resulting in their partials exhibiting important amplitude and frequency variations. The Fourier transform (FT), or, in practice, the fast Fourier transform (FFT), is commonly used to determine the frequency components of a given waveform but is not very well suited to nonstationary signals whose spectral content varies with time. In order to treat these signals, the short term Fourier transform (STFT) is usually employed. However, it is not exempt from drawbacks. There is, for example, the classic time–frequency resolution dilemma (Flandrin,[1] pp. 10–16). The wavelet transform (WT) or the nonlinear Wigner–Ville transform can partially overcome this problem, but they are not well suited for musical signals like piano sounds, because, in addition to the difficulty of separating frequencies of closely spaced frequency components for the former and to the presence of interference terms for the latter (Torrésani[2] and Hlawatsch[3]), they do not work well on a wide frequency range, as much as five octaves (Rossi *et al.*[4,5]). It must be noted that the number of partials can range from as few as eight usable partials for note B5 to 60 or more partials for note C1, while on a given length of signal their amplitudes can vary from 0 dB for the strongest partial taken as a reference, to −60 dB for the faintest usable one (see Fig. 2). Moreover, the number of significant partials depends greatly on the velocity of the piano hammer. This high number of frequency components obviously greatly increases if several notes are played simultaneously in the case of polyphony, and weak partials can then be very close (in frequency) to strong ones (Rossi[4,5]). In his book about wavelets, Torrésani[2] states (p. 87, 1.25-6) that ''when a signal is composed of several spectral components, the problem of their determination becomes more difficult if their frequencies are closely spaced'' (translated by the authors), which is the case here. In the same book (p. 89, 1.4-5), Torrésani notes that ''it is crucial to know in advance the values of the frequencies we want to find.'' Obviously, it is not possible to know in advance the number and the frequencies of the spectral peaks present in a given piano sound, even in the case of monophony. The problem is even more complicated by the fact that these components are not stationary, presenting great amplitude variations with time, and, for any given time, from partial to partial.

The Wigner–Ville Transform (WVT), a nonlinear transform, produces parasitic cross-terms whose number is $N(N-1)/2$ for N frequency components. The use of an analytic signal removes half of these unwanted cross-terms, and some transforms of Cohen's class (Choï–Williams, for example) can spread them in the time–frequency plane (Cohen,[6] pp. 126–127). However, it is worth noting that nowhere in their two reference books, neither Cohen[6] nor Flandrin[1] state how many decibels the parasitic terms can be reduced to. In other words, is the interference reduction efficient enough to allow the detection of nonstationary frequency components a few Hertz apart with an amplitude difference in the −40 to −50 dB range? It does not seem to be so, and this is probably why the STFT even now is still used extensively. As a consequence, most sound identification procedures simply use the STFT. They are usually based on the determination of the exact positions of partials' spectral peaks (Pizsczsalski,[7]

[a]Electronic mail: lrossi@univ-corse.fr

Pearson,[8] Doval,[9] and Watson[10]), although sometimes only the bin numbers corresponding to the position of the maximum induced by the partials are necessary (Brown[11] and Rossi *et al.*[4,5]). It is thus important to evaluate the possible variation of these spectral peaks for piano signals in order to see if the cited procedures are applicable to these kinds of signals and to quantify these variation phenomena in order to take them into account for the development of new identification procedures. In this paper we show that almost unpredictable changes of the position and/or the splitting of some STFT spectral lines corresponding to a given partial are mainly caused by changes in the instantaneous frequency (IF) of the partial.

II. NOTION OF INSTANTANEOUS FREQUENCY

The idea of frequency is very clear when applied to periodic signals: It represents the number of times a signal is identically reproduced during a unit of time, usually a second. For a simple sinusoid of the form $u(t)=a\cos(\omega t+\varphi)$, with a, ω, and φ constant, the frequency is $f=\omega/2\pi$. Even if the form of the signal is more complicated, similar definitions can be used, as long as all the parameters defining the signal remain constant. This is the case, for example, of a sawtooth waveform or a pulse train, with constant parameters. Their frequency is easy to determine, even if the form of the signal is far from sinusoidal.

A more subtle notion of instantaneous frequency (IF) is quite old and can be traced back to Helmholtz,[12] Carson,[13] and Gabor,[14] but Ville[15] can be considered the first to have defined the notion of an analytic signal and to have used it to derive the IF of a given signal. Boashash[16] gave a review of the notion of IF and of related problems, and it will thus be sufficient to recall here some definitions and properties of the analytic signal and of the IF.

Following Ville,[15] the analytic signal $z(t)$ associated with a real signal $u(t)$ is defined by

$$z(t)=u(t)+j\hat{u}(t), \tag{1}$$

where $\hat{u}(t)$ is

$$\hat{u}(t)=H[u(t)]=\text{p.v.}\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{u(\zeta)}{t-\zeta}d\zeta, \tag{2}$$

where p.v. stands for the Cauchy principal value; $\hat{u}(t)$ is thus defined as the Hilbert transform (HT) of $u(t)$, and signals $u(t)$ and $\hat{u}(t)$ are said to be in quadrature; $z(t)$ is a complex-valued function of time t and can be written as $z(t)=A(t)\exp[j\theta(t)]$. The IF and the envelope of $u(t)$ are then defined by the following: Instantaneous frequency [(IF) (Vakman[17])]:

$$f(t)=\frac{1}{2\pi}\frac{d}{dt}\left(\arctan\frac{\hat{u}(t)}{u(t)}\right)=\frac{1}{2\pi}\frac{\hat{u}(t)'u(t)-u(t)'\hat{u}(t)}{u(t)^2+\hat{u}(t)^2}; \tag{3}$$

amplitude envelope of $u(t)$:

$$A(t)=|z(t)|=\sqrt{u(t)^2+\hat{u}(t)^2}. \tag{4}$$

The angular frequency $[\omega(t)]$ can be considered as the speed (in radians/s) at which the vector $[u(t),\hat{u}(t)]$ rotates around the origin, while the IF is simply $f(t)=\omega(t)/2\pi$. This is a generalization of the behavior of the signal $u(t)=a\cos(\omega t+\varphi)$, associated with $\hat{u}(t)=a\sin(\omega t+\varphi)$.

Besides many useful properties, the Hilbert transform H, is characterized by the following facts:

$$H \text{ is a linear operation,} \tag{5}$$

$$H\{H[u(t)]\}=-u(t), \tag{6}$$

$$H[a\cos(\omega t+\varphi)]=a\sin(\omega t+\varphi), \tag{7}$$

if a, ω and φ are constant.

However, in general, $H[\cos(\Phi(t))]$ is different from $\sin[\Phi(t)]$.

Other properties apply to the spectra of $u(t)$, $\hat{u}(t)$, and $z(t)$, mainly: If $U(j\omega)$, $\hat{U}(j\omega)$, and $Z(j\omega)$ are FT of $u(t)$, $\hat{u}(t)$, and $z(t)$ respectively, then

$$\hat{U}(j\omega)=\begin{cases}-jU(j\omega), & \text{if } \omega>0\\ 0, & \text{if } \omega=0,\\ jU(j\omega), & \text{if } \omega<0\end{cases}$$

and thus

$$Z(j\omega)=\begin{cases}2U(j\omega), & \text{if } \omega>0,\\ U(j\omega), & \text{if } \omega=0,\\ 0, & \text{if } \omega<0.\end{cases} \tag{8}$$

These definitions are applicable to continuous-time systems, but the computations of complex signals were made on a computer in a MATLAB®[18] environment. The Hilbert routine included in MATLAB®[18] was used to compute the Hilbert transforms of the various signals. This routine first performs a FFT on signal $u(t)$, and then uses relations (8) and an inverse FFT to get the Hilbert transform $\hat{u}(t)$. The derivatives used in relation (3) were computed as $u(nT)'=[u(nT)-u((n-1)T)]/T$, $T=1/f_s$ being the sampling period with the same kind of relation for $\hat{u}'$. Taking the Z transform of both sides of this relation, one get the transfer function $H(z)=U'(z)/U(z)=(1-z^{-1})/T$. This transfer function computed at real frequencies, e.g., for $z=e^{j\omega t}$, gives $|H(j\omega)|=(2/T|\sin(\omega T/2)|=2f_s|\sin(\pi f/f_s)|$, while $\arg[H(j\omega)]=\pi/2-\omega T/2$. If f/f_s is small enough, then $\sin(\pi f/f_s)$ is close to $\pi f/f_s$ and $|H(j\omega)|\simeq 2\pi f=\omega$, with $\arg[H(j\omega)]\simeq\pi/2$.

If we take, for example, $f=1480$ Hz for the highest frequency used and $f_s=22\,050$ Hz, then $\pi f/f_s=0.210\,486$ while $\sin(\pi f/f_s)=0.209\,305$, so that the error is of the order of -0.7%, small enough to consider the above approximation as correct.

In order to understand how the IF can vary, it will first be computed on elementary signals.

A. Instantaneous frequency of elementary signals

1. Constant amplitude and frequency sinusoid

The formula $u(t)=a\cos(\omega t+\varphi)$ represents a sinusoid whose amplitude and frequency are constant if the parameters a, ω, and φ are constant. Relation (7) gives $\hat{u}(t)=a\sin(\omega t+\varphi)$, and, thus, $z(t)=ae^{j(\omega t+\varphi)}$, from which relations (3) and (4) give easily

$$|z(t)| = |a| \quad \text{and} \quad d\{\arg[z(t)]\}/dt = \omega.$$

These values of the amplitude and of the angular frequency are the ones expected, but things are quite different when considering two sinusoids.

2. Sum of two sinusoids

Suppose now that the signal considered is expressed as $u(t) = A_1 \cos(\omega_1 t) + A_2 \cos(\omega_2 t + \varphi)$, with

$$\begin{aligned} &\omega_1 = 2\pi f_1, \quad A_1 > A_2 > 0, \\ &\omega_2 = 2\pi f_2, \quad f_1 > f_2, \end{aligned} \tag{9}$$

where A_1, A_2, f_1, f_2, and φ are constant, and f_1 and f_2 are the frequencies of the sinusoids. Due to properties (5) and (7) of the Hilbert transform,

$$z(t) = u(t) + j\hat{u}(t) = A_1 \cos(\omega_1 t) + A_2 \cos(\omega_2 t + \varphi) + j[A_1 \sin(\omega_1 t) + A_2 \sin(\omega_2 t + \varphi)],$$

and relations (3) and (4) give

$$|z(t)| = \sqrt{A_1^2 + A_2^2 + 2A_1 A_2 \cos[(\omega_1 - \omega_2)t + \varphi)]} \tag{10}$$

(amplitude envelope), and

$$f(t) = \frac{A_1^2 f_1 + A_2^2 f_2 + A_1 A_2 (f_1 + f_2) \cos[(\omega_1 - \omega_2)t + \varphi]}{A_1^2 + A_2^2 + 2A_1 A_2 \cos[(\omega_1 - \omega_2)t + \varphi]} \tag{11}$$

(instantaneous frequency).

Due to condition $A_1 > A_2 > 0$, the IF is always finite since the denominator of $f(t)$ is always strictly positive. Also, the expressions for the amplitude and the IF get simpler when $\cos[(\omega_1 - \omega_2)t + \varphi] = \pm 1$:

if

$$\begin{aligned} &\cos[(\omega_1 - \omega_2)t + \varphi] = 1, \quad |z(t)| = A_1 + A_2, \\ &f_{\max} = \frac{f_1 A_1 + f_2 A_2}{A_1 + A_2}, \end{aligned} \tag{12}$$

if

$$\begin{aligned} &\cos[(\omega_1 - \omega_2)t + \varphi] = -1, \quad |z(t)| = A_1 - A_2, \\ &f_{\min} = \frac{f_1 A_1 - f_2 A_2}{A_1 - A_2}. \end{aligned} \tag{13}$$

These values correspond, respectively, to the maximum and the minimum of the amplitude and of the instantaneous frequency. Note that these results could have been found without using the Hilbert transform and the analytic signal, and were known by Helmholtz.[12] The Gabor[14]/Ville[15] method used to determine the IF of a sinusoid also applies for signals different from sinusoids (see the Cauchy function in Ville[15]), but it must be said that it is best suited to "narrow band" signals.

It is now time to consider some numerical examples. Taking

$$\begin{aligned} &A_1 = 1, \quad f_1 = 1000\ \text{Hz}, \\ &A_2 = 0.9, \quad f_2 = 1005\ \text{Hz}, \quad \varphi = \pi/4. \end{aligned}$$

The above relations immediately give the following:

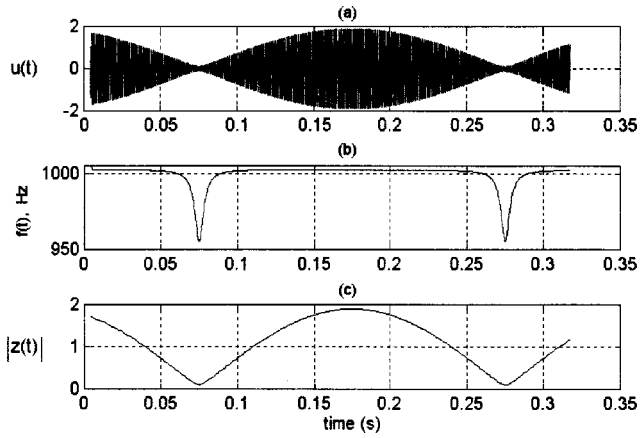


FIG. 1. (a) Signal, $u(t) = A_1 \cos(2\pi f_1 t) + A_2 \cos(2\pi f_2 t + \varphi)$, with $A_1 = 1$; $A_2 = 0.9$; $f_1 = 1000$ Hz; $f_2 = 1005$ Hz; $\varphi = \pi/4$. (b) Instantaneous frequency. (c) Amplitude.

maximum of the envelope: $E_{\text{man}} = A_1 + A_2 = 1.9$,

minimum of the envelope: $E_{\min} = A_1 - A_2 = 0.1$,

maximum of the IF: $f_{\max} = 1002.4$ Hz,

minimum of the IF: $f_{\min} = 955$ Hz.

The upper part of Fig. 1 shows $u(t)$, while the middle part shows the IF and the lower part the envelope. The beats between the two components of $u(t)$ are clearly visible, and the time between the two minima of the envelope is, as expected, the inverse of the difference in frequency of the two sinusoids, i.e., $T = \frac{1}{5} = 0.2$ s. Note that the amplitude computed with the help of the Hilbert transform represents correctly the envelope of $u(t)$. It is also clear that (1) the IF reaches the $f_{\max}$ and $f_{\min}$ bounds, (2) the IF is most of the time close to f_1 and f_2, and (3) the differences between f_1 and $f_{\min}$, and f_2 and $f_{\min}$ can be quite great: 45 and 50 Hz here. This difference is a function of A_1, A_2, f_1, and f_2, and increases if A_2 gets closer to A_1, and/or if f_2 becomes more distant from f_1. With the same values of f_1 and f_2 and with the value $A_2 = 0.95$, $f_{\min}$ becomes equal to 905 Hz, while with values $A_2 = 0.9$ and $f_2 = 1007$ Hz, we have $f_{\min} = 937$ Hz. Other conditions like A_1 and/or $A_2 < 0$, or $f_2 < f_1$ can be treated the same way without any problems.

The important thing to note here is that even with individual frequencies f_1 and f_2 constant and close to each other, depending on the values of f_1, f_2, A_1, and A_2, the IF of the sum of the two sinusoids can be quite different from f_1 and f_2, and generally presents large variations under or above these frequencies. This kind of analysis can be extended to more than just two sinusoids, and it is now possible to understand easily what happens to the IF of partials from piano notes.

III. APPLICATION OF THE INSTANTANEOUS FREQUENCY TO PIANO PARTIALS

Figure 2 shows the temporal evolution of note B3 of a piano sampled with a frequency of 22 050 Hz and its spectrum computed with a 32 768-point FFT. The sixth partial of this note was then bandpass filtered with a ± 4.7 Hz at -3

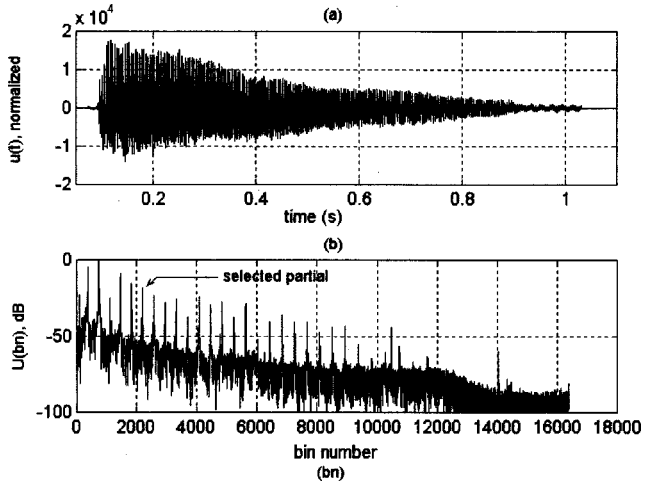


FIG. 2. (a) Temporal evolution of note B3 of a piano. (b) Spectrum of note B3.

dB (± 15.5 Hz at -40 dB) Gaussian filter, and an inverse FFT was performed to return it to the temporal domain. The results of these operations are visible in Fig. 3, which shows the spectrum filtered around the position of the sixth partial and the temporal evolution of the filtered partial. A fast rise of the amplitude is visible at the beginning, and then there is a slow decay with beats due mainly to a slight mistuning among the three strings used for note B3. Weinreich[19] explained that this phenomenon is due to the interaction among the partials produced by each string struck and showed that they occur almost inevitably, even with well-tuned pianos.

The noncausality of the filter used is visible when comparing Figs. 2 and 3. On the latter figure, the partial begins at time 0 s instead of approximately 0.1 s, as in Fig. 2. This is an artifact of the method used to recover the temporal evolution of the partial and does not change the overall conclusions.

The variations of the amplitude (in dB) and of the IF (in Hz, computed with the help of the HT) are given in Fig. 4. The IF is quite stable and close to 1479 Hz, except at times $t_1 = 0.61$ s and $t_2 = 1.05$ s, when the amplitude is low; at

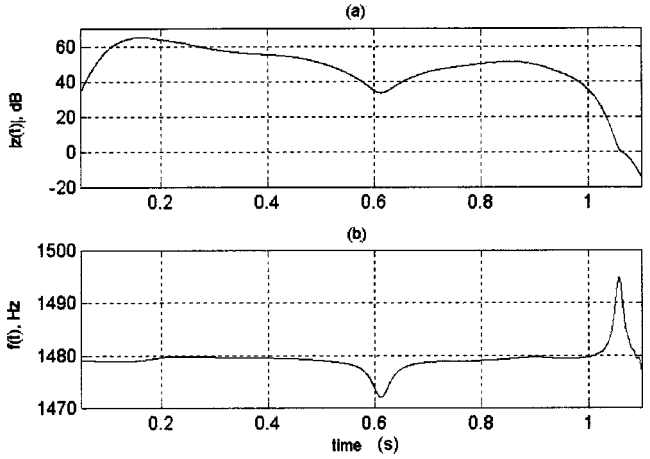


FIG. 4. (a) Amplitude of partial #6 of note B3. (b) Instantaneous frequency.

these instants this IF shows important variations, -7 Hz (-0.5%) and $+15$ Hz ($+1\%$), respectively. The evolution of this IF can be explained as follows: Suppose that three components are present in the signal with frequencies f_1, f_2, f_3, amplitudes A_1, A_2, A_3, and zero initial phase due to the synchronous excitation of the three strings by the hammer. Choosing $f_1 = 1479$ Hz, $A_1 = 1$, $f_2 = 1479.6$, $A_2 = 1.35$, and $f_3 = 1480.23$ Hz, $A_3 = 0.92$, gives Fig. 5 which shows variations of the IF very close to those of Fig. 4. Note that these values were found by a method of successive approximations. The cut and try method used is not detailed here since the goal was simply to give a set of frequencies and amplitudes able to reproduce the signal, and not to give a method to determine these values. However, they could have been determined by high-resolution methods like the one described by Laroche.[20]

The difference in frequency between these components is very small, 1.23 Hz at the most, which, when translated to the fundamental of note B3, is equivalent to a difference of 0.21 Hz. Since this corresponds to a relative variation of 0.014% (one cent is about 0.06%), the piano can be considered to be very well tuned, at least for note B3. Of course, this difference is much less than what the human ear can

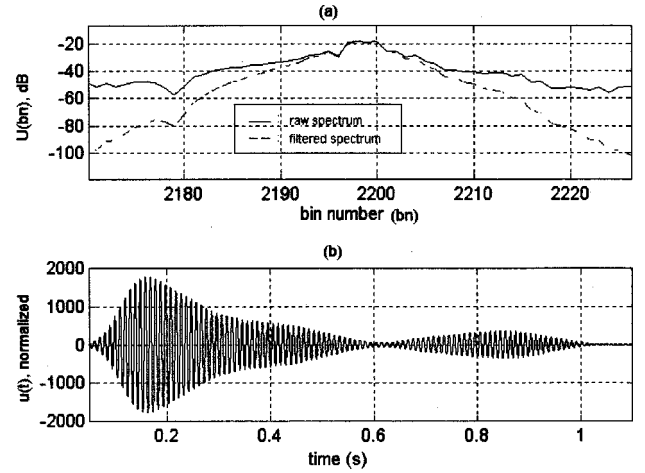


FIG. 3. (a) Raw and filtered spectrum of partial #6 of note B3 of a piano. (b) Temporal evolution of the partial filtered by a Gaussian band-pass filter [bandwith of ± 4.7 Hz at -3 dB (± 18 Hz at -50 dB)].

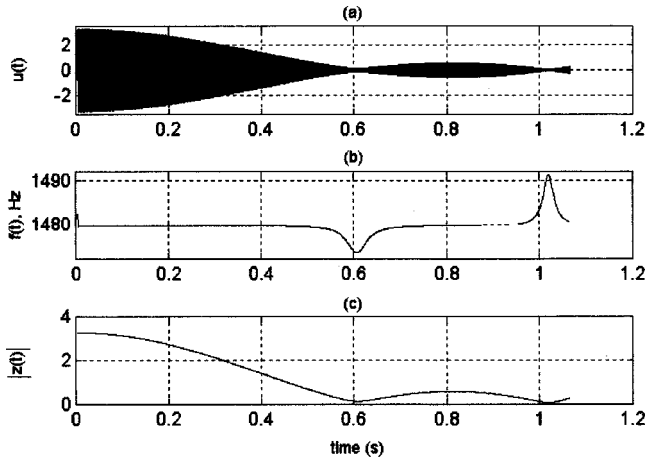


FIG. 5. (a) Signal $u(t) = A_1 \cos(2\pi f_1 t) + A_2 \cos(2\pi f_2 t) + A_3 \cos(2\pi f_3 t)$, with $A_1 = 1$; $A_2 = 1.35$; $A_3 = 0.92$ and $f_1 = 1749$ Hz; $f_2 = 1479.6$ Hz and $f_3 = 1480.23$ Hz. (b) Instantaneous frequency. (c) Amplitude.

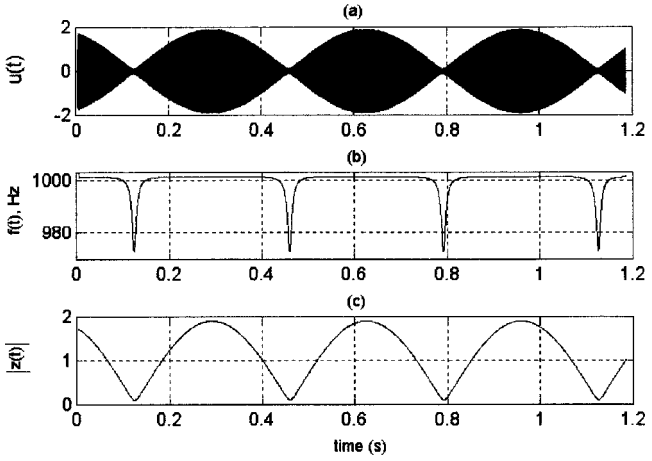


FIG. 6. (a) Signal $s(t)=\cos(2\pi f_1 t)+0.9\cos(2\pi f_2 t+\pi/4)$, with $f_1=1000$ Hz; $f_2=1003$ Hz. (b) Instantaneous frequency. (c) Amplitude.

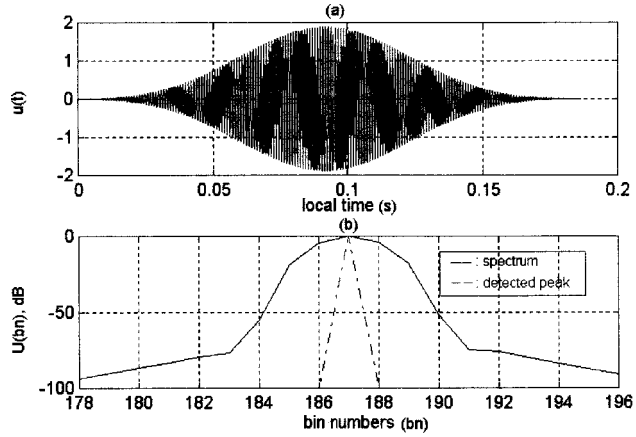


FIG. 7. (a) Signal $s'(t)$:$s(t)=\cos(2\pi f_1 t)+0.9\cos(2\pi f_2 t+\pi/4)$ (with $f_1=1000$ Hz; $f_2=1003$ Hz) weighted by a Blackman window positioned midway between two minimums of $s(t)$. (b) Spectrum of $s'(t)$: only one peak is detected.

discriminate, but the point is that when a STFT is performed on this signal one will see a splitting, for a few particular moments, of the main spectral peak into one or two adjacent peaks, and this, in turn, could be enough to perturb a note identification process. The problem is to recognize that at some instants the spectral peaks split, that this phenomenon is tied to changes in the IF of the signal at these instants, and that it is unavoidable and, thus, must be treated as such.

Three points must be made clear here. In the first place, the bandwidth used to perform the bandpass filtering is by no means critical and could be increased to at least ± 10 Hz without changing the general conclusions. Second, a Gaussian filter was used because this kind of filter is very easy to implement, has good roll-off characteristics, and minimizes the $\Delta F \Delta t$ product, thus optimizing the time–frequency incertitude. Finally, the note (B3) and the partial (#6) chosen are by no means special; they can be considered as corresponding to the general case, i.e., due to the presence of three strings for each note (except, of course, for the lowest octaves), the IFs of the partials of almost all the notes of the piano exhibit this kind of fluctuation. This problem is furthermore compounded in the case of polyphony, because two different notes can have partials with very close frequencies, and this, in turn, will give beats among their components (Maher[21]). As a general rule, one can expect important variations on the amplitude and on the IF of the partials in a polyphonic context.

A. Variations of the instantaneous frequency and positions of the peaks of the STFT of test signals

In order to evaluate the consequences of the fluctuations of the IF of a signal on the position of the peaks of a STFT applied to the same signal, the following test signal was used:

$$s(t)=A_1\cos(2\pi f_1 t)+A_2\cos\left(2\pi f_2 t+\frac{\pi}{4}\right), \tag{14}$$

with

$$f_1=1000\,\text{Hz}, \quad f_2=1003\,\text{Hz}, \quad A_1=1, \quad A_2=0.9.$$

Formulas (12) and (13) give $f_{\text{minimum}}=973$ Hz, and $f_{\text{maximum}}=1001.4$ Hz.

Figure 6 shows the temporal variations of $s(t)$, of its IF, and of its amplitude envelope. In the following figures, the signal, sampled at 22 050 Hz, was analyzed with a $N=4096$ point (186 ms) STFT, a Blackman window [described by the relation $W(n)=0.42+0.5\cos[(2\pi/N)n]+0.08\cos[(2\pi/N)2n]$ with $n=-(N-1)/2,\ldots,-1,0,1,\ldots,\overline{N/2}$], and a hop interval of 20 ms (corresponding to a window advance of 441 points for each frame). The Blackman window (Harris[22]) was chosen because the associated sidelobes are about -60 dB under the level of the main lobe, and this, in turn, insures that no parasitic peaks above this level would be created by this artifact of the FFT. The price to pay for this low level of peaks is an increase in the main lobe width, but some note identification processes can overcome this problem easily (Rossi *et al.*[4,5]). The frequency difference between consecutive bins is $\Delta f=22\,050/4096=5.38$ Hz, and each bin lobe is ± 4.52 Hz at -3 dB wide. (Below, brackets [ ] mean that the value between brackets is rounded to the nearest integer.) With the given values of f_1 and f_2, and the 5.38 Hz frequency resolution, peaks at positions $1+[1000/5.38]=187$, and $1+[1003/5.38]=187$ contribute to the spectrum (1 is added because bin #1 corresponds to the dc part of the spectrum). The two components of the signal share the same bin, and that the same situation occurs when the IF is close to these frequencies. However, when the IF is close to 973 Hz, it will be associated with bin #1 $+[973/5.38]=182$, and thus a peak at this position can be expected.

Figure 7 gives result for the case when the analysis window is positioned midway between two minima of the signal: The spectrum is very regular and there is only one peak at position 187 (the triangular curve), which is detected by looking for the local maximum of the spectrum. Contrary to this, Fig. 8 shows a quite different situation when the analysis window is centered on a minimum of the signal, and, thus, on a minimum of the IF, because three peaks are now clearly visible at positions #188, #186, and #183. The first two correspond to the peak in position #187 of Fig. 7, which

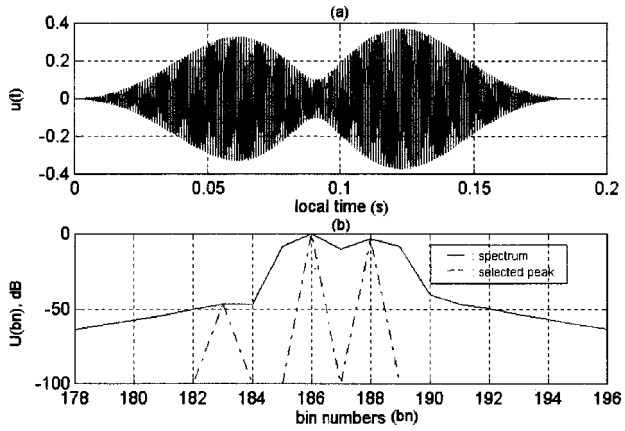


FIG. 8. (a) Signal $s'(t)$:$s(t)=\cos(2\pi f_1 t)+0.9\cos(2\pi f_2 t+\pi/4)$ (with $f_1=1000$ Hz; $f_2=1003$ Hz) weighted by a Blackman window centered on a minimum of $s(t)$. (b) Spectrum of $s'(t)$: three peaks are detected.

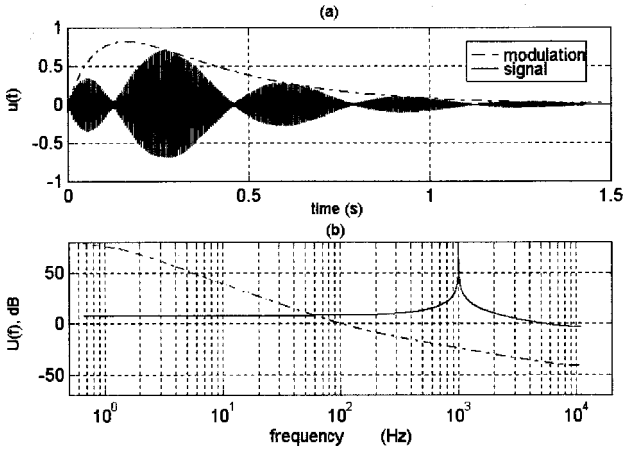


FIG. 10. (a) Signal $s'(t)=s(t).A(t)$, with $A(t)=\exp(-3.15t)-\exp(-11.025t)$ and $s(t)=\cos(2\pi f_1 t)+0.9\cos(2\pi f_2 t+\pi/4)$ with $f_1=1000$ Hz; $f_2=1003$ Hz. (b) Spectra of $s(t)$ and $A(t)$.

has split into two components because the signal analyzed by the FFT is then amplitude modulated. The third one corresponds to the minimum of the IF expected in bin #182. The peak is detected in position 183 instead of position 182 because the IF is close to its minimum for a very short time only, and because the FFT "averages" the frequencies. The peak associated with the minimum frequency is "pulled" slightly toward higher frequencies, and appears at position #183 instead of bin # 182.

The presence or absence of spectral components, as a function of time, is carried out by using successive 4096 point wide 20 ms shifted portions of the signal called frames. In the spectrum corresponding to each analyzed frame, local maxima indicate the presence of peaks. As a consequence, a spectral peak detection is realized for each frame, and gives the temporal evolution (presence or absence) of the signal's components. To plot this evolution (resulting from successive detections on the shifted frames) a graph is used, the *X* axis representing the time, and the *Y* axis the possible FFT bin numbers, the value of a bin number being increased by 0.5 when a peak is detected at that bin at a given instant. Figure 9 gives the results of this peak detection on the signal $s(t)$. A spectral peak corresponding to bin #187 is always present, except for short periods when the amplitude of the signal $s(t)$ is very low. When this happens, the IF is close to its minimum, and the peak in position #187 splits into two peaks at positions #186 and #188, while a peak at position #183 appears during one or two frames. Note also that the instants at which the IF is at its minimum (or maximum) are shifted slightly in the frame-by-frame procedure (compare Figs. 6 and 9) because at least one-fourth of the data in the FFT window must change before the spectrum shows significant differences. This remark obviously applies to the other figures as well.

In order to analyze a situation closer to piano sounds, the above signal $s(t)$ was multiplied by an amplitude function $A(t)$ giving the new test signal, $s'(t)=s(t)\cdot A(t)$, with

$$A(t)=\exp(-3.15t)-\exp(-11.025t),$$

and thus,

$$s'(t)=[\cos(2\pi f_1 t)+0.9\cos(2\pi f_2 t+\pi/4)]\cdot[\exp(-3.15t)-\exp(-11.025t)].$$

The maximum of the $A(t)$ occurs at time $t_{\max}=0.159$ s. The temporal evolution of the signal $s'(t)$, its modulation, and the spectra of $s(t)$ and $A(t)$, are shown in

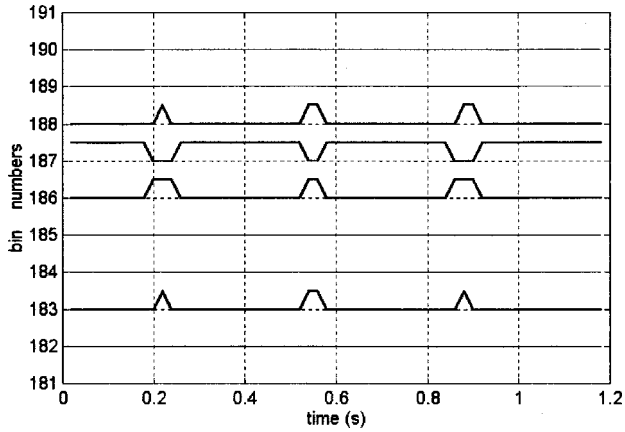


FIG. 9. Peak detection diagram for signal $s(t)=\cos(2\pi f_1 t)+0.9\cos(2\pi f_2 t+\pi/4)$ with $f_1=1000$ Hz; $f_2=1003$ Hz.

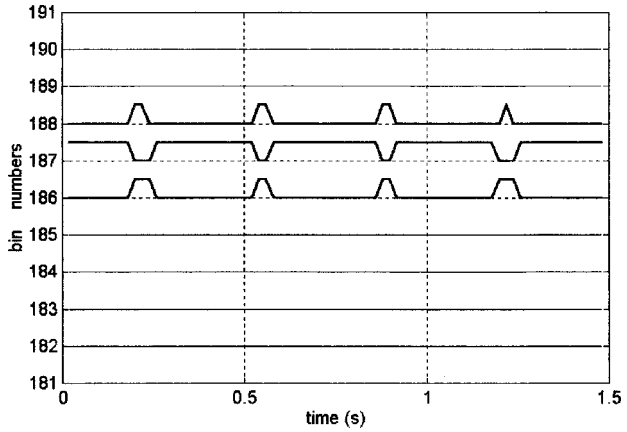


FIG. 11. The peak detection diagram for signal $s'(t)=s(t)\cdot A(t)$, with $A(t)=\exp(-3.15t)-\exp(-11.025t)$ with $s(t)=\cos(2\pi f_1 t)+0.9\cos(2\pi f_2 t+\pi/4)$, with $f_1=1000$ Hz; $f_2=1003$ Hz.

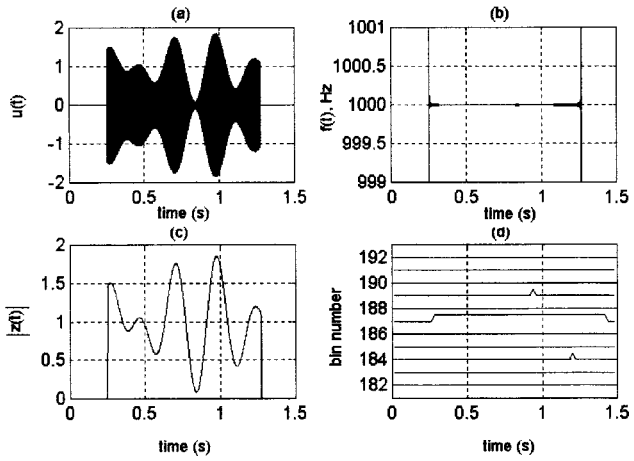


FIG. 12. (a) Signal $s_1(t)=\cos(2\pi f_0 t)\cdot[1+0.46\cos(2\pi f_1 t)+\cos(2\pi f_2 t)]$, with $f_0=1000$ Hz; $f_1=3$ Hz; $f_2=4.15$ Hz. (b) Instantaneous frequency. (c) Amplitude. (d) Frame-by-frame detection of the components of $s_1(t)$.

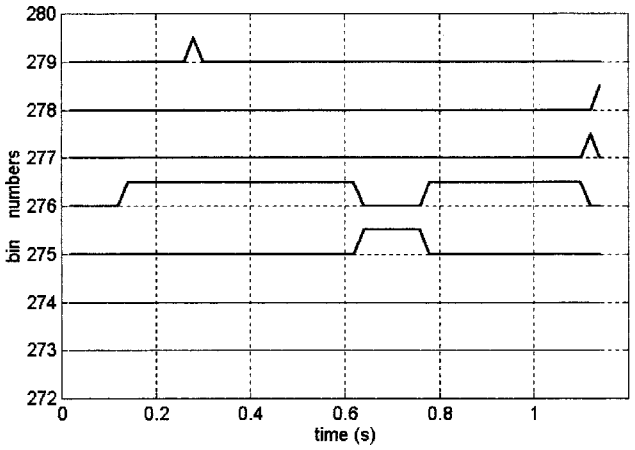


FIG. 13. The peak detection of frequencies around 1479 Hz (partial #6) for note B3 (4096-point Blackman window).

Fig. 10, while the result of the peak detection performed on signal $s'(t)$ is shown in Fig. 11. Note that due to the very low-frequency components of $A(t)$, there are no differences between the spectra of $s'(t)$ and $s(t)$. The values of the time constants in $A(t)$ were chosen to give a realistic rendition of a piano's amplitude versus time envelope.

The results shown in Figs. 9 and 11 do not differ dramatically, and that conclusions are similar, except that the line corresponding to the minimum of the IF (bin #183), is not visible in Fig. 11.

From what precedes, another question arises: Are the phenomena described tied to the changes of the IF, to the changes in the envelope of the signal, or possibly to both? Figure 12 gives a partial answer. The top left figure represents $s_1(t)$, a 1000 Hz sinusoid whose amplitude is modulated by the sum of two sinusoids, i.e.,

$$s_1(t)=\cos(2\pi f_0 t)\cdot[1+0.46\cos(2\pi f_1 t)+\cos(2\pi f_2 t)],$$

with: $f_0=1000$ Hz, $f_1=3$ Hz, and $f_2=4.15$ Hz.

As expected (following Ville[15]), the IF of $s_1(t)$ computed as explained in Sec. I A is constant and equal to $f_0=1000$ Hz, as can be seen in the top right figure. The bottom right figure shows the frame-by-frame detected spectra. It appears that line #187 is present during the complete duration of the signal without splitting into adjacent lines observed when the IF varies. Two very short lived detection peaks appear on lines #184 and #189, and correspond to the two deep minima around time 1 s. These two peaks are slightly delayed due to the STFT as already stated in Sec. I B 4. The problem is that when the analysis window covers a minimum like the one shown in Fig. 8, the signal is amplitude modulated, and the FFT produces a few spurious components. This problem is much less important than the splitting of the lines already described. In fact, these spurious components do not appear systematically at each minimum, have very short duration, and would be rarely noticeable on a real signal due to inherent noise.

The upshot of this discussion is that both the IF and the amplitude envelope minima contribute to the changes in the spectra, but the IF is dominant in the process.

B. Application of the instantaneous frequency to a piano note

The same procedure was applied to a real sound, namely the note B3 already treated in Sec. I B C. Figure 4 gives the IF of partial #6 whose nominal value is, as already stated, 1479 Hz, with a minimum frequency of 1472 Hz at time $t_1=0.61$ s, and a maximum frequency of 1495 Hz at $t_2=1.05$ s. With these values of the IF, a spectral peak is expected most of the time at position $1+[1479/5.38]=276$, except perhaps at time t_1, where peaks at positions 276 ± 1, and $1+[1472/5.38]=274$, can be expected, and at time t_2 where peaks located at positions 276 ± 1, and $1+[1495/5.38]=279$ can also be expected. Figure 13 gives the results of the frame-by-frame peak detection, and apart from the transitory peak at position #279 (time $=0.3$ s), and the absence of a peak at position #274, the expected peaks are present, slightly delayed, as explained above. The peak at position #276 is always visible, except at the very beginning, when the analyzed signal is too weak, and around times 0.7 and 1.1 s, corresponding, respectively, to the times of the minimum and of the maximum of the IF. As already seen in Sec. I B 4, the peak in position #276 disappears at these moments and is replaced by one or two peaks in adjacent positions.

At time 0.7 s, the peak in position 276 is replaced by a peak in position 275, and this explains why no peak is visible in position 274. As already seen in Fig. 8, the peak corresponding to the minimum of the IF is pulled to a slightly higher position and is thus merged with the peak in position 275. The same kind of situation occurs at time 1.1 s, where the IF shows a maximum. The value of this maximum is high enough (1495 instead of 1479 Hz), to make the peak in position 276 disappear and be replaced by the peaks in successive positions 277 and 278. Note that the peak expected in position 279 is pulled down a little lower by the averaging effect of the FFT. Apart from a few insignificant false detections around time 0.4 s on position #273, changing to the Hamming window gives very similar results, as can be seen in Fig. 14. The same kind of phenomena, i.e., the splitting of a spectral line and/or the appearance of some new lines, occurs every time the IF shows significant departures from its

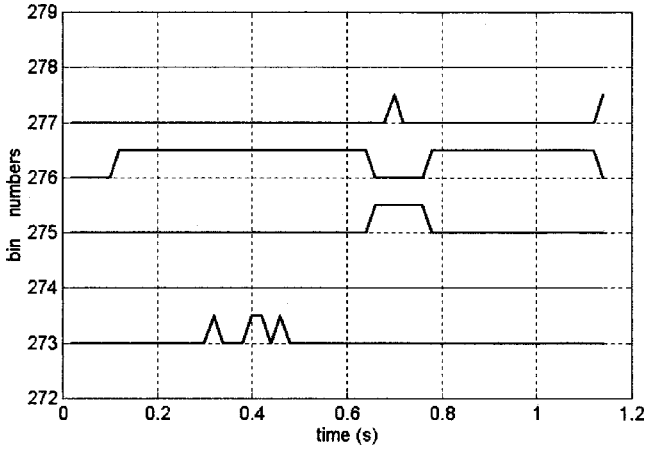


FIG. 14. The peak detection of frequencies around 1479 Hz (partial #6) for note B3; 4096-point Hamming window.

nominal frequency, regardless of the weighting window chosen, and it is thus necessary to take this situation into account in any identification procedure.

1. Use of shorter windows

The previous analysis was based on a 4096-point Blackman window and a sampling frequency of 22 050 Hz, and, thus, on a 186 ms frame length. This time may seem too long for piano signals, since notes as short as 125 ms (1/32nd note with a tempo of one second for a quarter-note) can be commonly found. However, the problem is actually not very serious because the partials of a signal are correctly detected as soon as this signal fills in between one-fourth (about 46 ms) and one-third (about 60 ms) of the frame (Rossi *et al.*[4,5]). It is, nevertheless, interesting to look at the case of shorter frames, such as the 1024-point FFT frame case in the following example.

Figure 15 gives the results of the identification with 1024-point FFT frames. With this window width, partial #6 of note B3 is expected on bin $\#1+[1479/(5.38*4)]=70$. Apart from the two false detections at the very beginning, this is clearly the case, except again around the instant 0.6 s, where the position of the detected line shifts to bin #69, and at the end where bin #71 is excited. These results, which, as already stated, are by no means specific to note B3, show that important changes can be expected in the position of detected spectral lines every time the IF presents important variations.

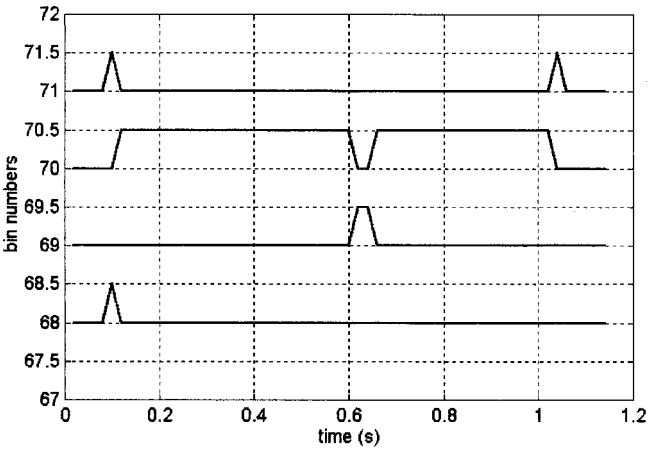


FIG. 15. The identification of partial #6 of note B3 with a 1024-point Blackman window.

Shorter windows would give proportionally wider FFT filters, and thus the variations of the IF would become less visible in the peak detection diagram. It is nevertheless clear that the shorter the window, the lower the frequency resolution, and the more difficult it will be to identify notes within octaves 1, 2, and 3, corresponding to notes C1 to B3.

The 4096- (186 ms) point window seems to be a good compromise between time and frequency resolution because, as already stated, it is not necessary for the signal to fill the FFT window in completely to detect its partials, and, furthermore, such a window allows the separation of partials of octave 1 and above.[4,5]

IV. CONCLUSIONS

In this paper we point out two important phenomena that must be taken into account by any method for the identification of the fundamental frequency of musical sounds. The first one is that the instantaneous frequency of a piano note partial can present important variations due to the beats occurring when several partials are closely spaced in the frequency domain. This is the case for the partials produced when the two or three strings associated with a piano note are struck and also, very often, in polyphonic sounds. This result implies that, for these sounds, the use of the temporal evolution of the frequency of partials for the procedure of note identification is impossible. The second important phenomenon is that the instantaneous frequency variations of a given partial can consequently modify the detection of the STFT spectral peaks associated with the partial because these peaks can temporarily disappear or be replaced by adjacent peaks. This imposes that any note identification algorithm must integrate the possible temporary absence of STFT spectral lines and the temporary presence of adjacent components.

References 4 and 5 describe a polyphonic piano note identification method that takes these problems into account and whose rate of correct identification reaches up to 92% for up to four notes played simultaneously on the same piano. This high success rate is partly due to the fact that both the possible absence and the splitting of the spectral lines were treated as being an inevitable part of the process. The above results also show that depending on the goal to reach, it might not be desirable to look for too high a frequency resolution. This was the case in the note identification procedure described by Rossi *et al.*,[4] but obviously, it would be very interesting to follow the frequency variations in the case of sound synthesis.

[1] P. Flandrin, *Temps–Fréquence* (Hermès, Paris, 1993).
[2] B. Torrésani, *Analyse Continue Pour Ondelettes* (InterEditions/CNRS Editions, City, 1995).
[3] F. Hlawatsch, "Interferences terms in the Wigner distribution," *International Conference on Digital Signal Proceedings*, Florence, 1984, pp. 363–367.
[4] L. Rossi, G. Girolami, and M. Leca, "Identification of polyphonic piano sounds," Acta Acust. **83**, 1077–1084 (1998).

[5] L. Rossi, ''Identification de sons polyphoniques de piano,'' Ph.D. thesis, University of Corsica, 1998.
[6] L. Cohen, *Time–Frequency Analysis* (Prentice–Hall, Englewood Cliffs, NJ, 1995).
[7] M. Piszczalski and A. Galler, ''Predicting musical pitch from component frequency ratios,'' J. Acoust. Soc. Am. **66**, 710–720 (1979).
[8] E. R. S. Pearson, ''The multiresolution Fourier transform and its application to polyphonic audio analysis,'' Ph.D. thesis, University of Warwick, 1991.
[9] B. Doval, ''Estimation de la fréquence fondamentale des signaux sonores,'' Ph.D. thesis, University of Paris VI, 1994.
[10] C. R. Watson, ''The computer analysis of polyphonic music,'' Ph.D. thesis, University of Sydney, 1985.
[11] J. C. Brown, ''Musical fundamental frequency tracking using a pattern recognition method,'' J. Acoust. Soc. Am. **92**, 1394–1402 (1992).
[12] X. Helmholtz, *On the Sensations of Tone*, 2nd ed. (Dover, New York, 1954), pp. 414–415.
[13] J. R. Carson and T. C. Fry, ''Variable frequency electric circuit theory,'' Bell Syst. Tech. J. **16**, 513–540 (1937).
[14] D. Gabor, ''Theory of communication,'' J. Inst. Electr. Eng., Part 1 **903**, 429–457 (1946).
[15] J. Ville, ''Théorie et applications de la notion de signal anamytique,'' Câbles Transm. **1**, 61–74 (1948).
[16] B. Boashash, ''Estimating and interpreting the instantaneous frequency of a signal—Part 1: Fundamentals,'' Proc. IEEE **80**, 520–538 (1992).
[17] D. YE. Vakman, ''On the definition of concepts of amplitude, phase and instantaneous frequency of a signal,'' Radio Eng. Electron. Phys. **17**, 754–759 (1972).
[18] MATLAB, version 5.3, Mathworks product, 1999.
[19] G. Weinreich, ''Coupled piano strings,'' J. Acoust. Soc. Am. **62**, 1474–1484 (1977).
[20] J. Laroche, ''The use of matrix pencil method for the spectrum analysis of musical signals,'' J. Acoust. Soc. Am. **94**, 1958–1965 (1993).
[21] R. C. Maher, ''Evaluation of a method for separating digitized duet signals,'' J. Audio Eng. Soc. **38**, 956–979 (1990).
[22] F. J. Harris, ''On the use of windows for harmonic analysis with the Discrete Fourier Transform,'' Proc. IEEE **66**, 51–83 (1978).

On representing signals using only timing information

Ramdas Kumaresan[a] and Yadong Wang
Department of Electrical Engineering, Kelley Hall, 4 East Alumni Avenue, University of Rhode Island, Kingston, Rhode Island 02881

(Received 7 December 1999; revised 20 July 2001; accepted 1 August 2001)

It is well known that only a special class of bandpass signals, called real-zero (RZ) signals can be uniquely represented (up to a scale factor) by their zero crossings, i.e., the time instants at which the signals change their sign. However, it is possible to invertibly map arbitrary bandpass signals into RZ signals, thereby, implicitly represent the bandpass signal using the mapped RZ signal's zero crossings. This mapping is known as real-zero conversion (RZC). In this paper a class of novel signal-adaptive RZC algorithms is proposed. Specifically, algorithms that are analogs of well-known adaptive filtering methods to convert an arbitrary bandpass signal into other signals, whose zero crossings contain sufficient information to represent the bandpass signal's phase and envelope are presented. Since the proposed zero crossings are *not those of the original signal*, but only indirectly related to it, they are called hidden or covert zero crossings (CoZeCs). The CoZeCs-based representations are developed first for analytic signals, and then extended to real-valued signals. Finally, the proposed algorithms are used to represent synthetic signals and speech signals processed through an analysis filter bank, and it is shown that they can be reconstructed given the CoZeCs. This signal representation has potential in many speech applications. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1405523]

PACS numbers: 43.60.Lq, 43.72.Ar, 43.64.Bt [JCB]

I. INTRODUCTION

A key issue in sampling theory is the construction of a sequence of samples that unambiguously represent a signal $s(t)$. There are two major approaches to constructing such a sequence of samples.[1]

(1) The first is the familiar "Shannon sampling,"[2] i.e., define samples $\{s_n\}$ as the values taken by $s(t)$ on a given set of sampling points, $\{t_n\}$, i.e., $s_n = s(t_n)$.

(2) The second approach is the less familiar notion of representing signals by certain time instants..., $\tau_{-1}, \tau_0, \tau_1, \tau_2, \ldots$. Specifically, for example, in certain cases the zero-crossings or level-crossing locations of $s(t)$ can be used to represent $s(t)$ to within a scale factor.

Reconstruction is the process of the pointwise recovery of $s(t)$ given the sampling sequence. In this paper we are primarily concerned with the second approach, i.e., representing bandpass signals by certain time instants. However, in the proposed signal representation schemes, these timing instants are not the zero-crossing locations of $s(t)$ themselves (as in, for example, Refs. 3, 4), but the zero-crossing locations of certain functions that are related to the phase and envelope of the signal $s(t)$.

A motivation for signal representation based on timing instants comes from our desire to understand certain aspects of biological signal processing. The cochlea, or inner ear, is known to decompose an acoustic stimulus into frequency components along the length of the basilar membrane. This phenomenon is called a *tonotopic* decomposition.[5] Further, it is also known that the nerve fibers emanating from the cochlea convey information to the brain in the form of trains of almost identically shaped nerve impulses or spikes. Since the nerve spikes are all essentially stereotyped waveforms, the information about the acoustic signal is carried in the timing information contained in these spike trains in the form of spike rate or interspike intervals. See, for example, Refs. 6–8. This fact raises a fundamental question: can signals (except for a scale factor) be represented by timing information alone? Our purpose in this paper is to affirmatively answer this question and in the process propose a phenomenological model for signal processing by the cochlea. A model of such a signal processing system is shown in Fig. 1. We emphasize that this model is not intended to be an accurate physiological model of the inner ear but just a means to show that signals can be represented almost entirely by timing information. In this paper we are primarily concerned with the real-zero conversion (RZC) box in Fig. 1.

A. Previous related research

Bandpass signals whose zeros are distinct and real are called real-zero (RZ) signals.[1] They are described fully (up to a scale factor) by their zero crossings. For example, a sinusoid with frequency f_0 Hz is a RZ signal. It may be expressed in terms of its distinct zero crossings (located along the time axis at integer multiples of $1/2\,f_0$) as $\sin 2\pi f_0 t = Ct\Pi_{-\infty<k<\infty, k\neq 0}(1-2tf_0/k)e^{2tf_0/k}$. This is known as the Hadamard factorization[9] of a sinusoid. Extending this observation one step further, Voelcker and Requicha[10] raised the following interesting question: what are the conditions under which a bandpass signal, $s(t)$, might be recovered from its traditional zero crossings? Logan,[11] following up on their work, showed that a bandpass signal $s(t)$ can be represented by its traditional zero crossings in two special cases.

(1) If $s(t)$ and its Hilbert transform $\hat{s}(t)$, have no common zeros other than distinct real zeros and that the band-

[a]Electronic mail: kumar@ele.uri.edu

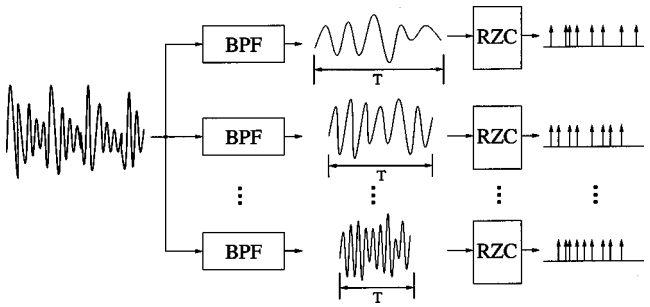


FIG. 1. Tonotopic real zero-crossing converter (RZC): The input signal is decomposed into bandpass signals by a set of bandpass filters. The bandpass signals are then viewed through an observation window of T seconds. Using a signal-adaptive algorithm over this time–frequency window, the signal is represented by a set of "covert zero crossings." (See the text for details.) This approach is motivated by the auditory periphery in which a composite signal is decomposed by a bank of frequency selective filters, and the information contained in the filtered signals is conveyed to the brain via timing information carried by nerve impulses.

width of $s(t)$ does not exceed an octave [one can get an intuitive understanding of this condition based on the following: If $s(t)$ and $\hat{s}(t)$ have a common zero, then the envelope itself goes to zero. The distinct real zero condition is required to ensure that the signal waveform has a zero crossing (i.e., a change of sign). Double real zeros will not produce a sign change in the waveform. The octave bandwidth constraint comes from the fact that the wider the bandwidth of a signal, the greater the number of zero crossings of the waveform are required to identify the signal uniquely, which, in turn, implies that $s(t)$ is a sufficiently high-frequency signal.] Logan provides a rigorous justification.

(2) If $s(t)$ is a full-carrier lower sideband (LSB) signal [i.e., $s(t)$ is a bandpass signal which has a large carrier at the high-frequency band edge]. Note that a full-carrier upper sideband signal may not have sufficient number of zero crossings to identify the signal uniquely.

Requicha,[1] in his lucid review paper, places Logans results in the general context of the theory of zeros of entire functions.

Although Logan's observations are interesting, there are two difficulties in using his results. First, they are existence theorems and do not provide a practical way to represent arbitrary bandpass signals by zero crossings or reconstruction algorithms. Second, most practical signals of interest like speech are time-varying signals, which need to be represented over short durations and hence Logan's theory based on strictly bandlimited signals is of limited use. Logan's final assessment in his paper is also pessimistic and states that "recovering a signal from its sign changes appears to be very difficult and impractical." [11]

In light of Logan's pessimism, researchers have attempted to find an invertible mapping that converts an arbitrary bandpass signal into a RZ signal; then one could use the zero crossings of the RZ signal to implicitly represent the bandpass signal. This process was dubbed "real zero conversion" (RZC) by Requicha.[1] This approach to the bandpass signal representation was investigated by Voelcker and his student, Haavik[12,1] and Bar-David.[13] Haavik[12,1] presented two transformations to accomplish RZ conversion: (1) repeated differentiation of the bandpass signal $s(t)$; and (2) the addition of a sine wave of known frequency equal to or higher than the highest frequency present in the signal and of sufficiently large amplitude, i.e., conversion of $s(t)$ into a full-carrier LSB signal. Zeevi and colleagues, in a series of insightful publications,[14,15] have extended the above ideas and applied them to one- and two-dimensional signals. Marvasti[16] and Hurt[9] have summarized and reviewed these ideas. Hurt[9] has compiled an extensive list of references related to zero crossings in one and two dimensions. Since the Fourier transform of a *time-limited signal* is the dual of a bandpass signal, many of the above results have counterparts in the frequency domain. This duality is explored in Ref. 15 (see also the references in Ref. 17). The above-mentioned RZC methods have practical drawbacks.[1] The repeated differentiation method is not very useful, because, differentiating a function more than a few times requires the use of extremely sharp filters to control the out-of-band noise. The sine wave addition method may introduce too many zero crossings than are needed to represent the bandpass signal.

In this paper we propose a novel signal-adaptive approach to RZC. Specifically we propose algorithms that are analogs of well-known adaptive filtering methods to convert $s(t)$ into other signals whose zero crossings contain sufficient information to represent the phase and envelope of $s(t)$. Since the zero crossings we advocate are *not those of the original signal* $s(t)$, we call them hidden or covert zero crossings (CoZeCs).

B. Organization of the paper and main results

The basic idea of our work is to try to represent signals by discrete time instants over short time intervals and frequency regions. The signals are confined to frequency regions by using a traditional filter bank. At the output of each filter, over a short duration, the envelope and phase of the signal is modeled using rational models. This is achieved by using an elegant signal adaptive algorithm called linear prediction in spectral domain (LPSD)[18] (Sec. IV). These rational models are then represented by certain zero crossings (CoZeCs), which then implicitly but essentially completely characterize the original signal. In effect, our results that exploit signal-adaptive methods are a significant extension of those due to Logan and Voelcker. Adaptive processing algorithms were not known or not yet prevalent during Logan and Voelcker's time (the 1960s and 1970s). The main results and the layout of the paper are as follows.

(i) *Modeling envelope and phase of bandpass signals:* In speech literature, the spectral envelope of a speech signal is traditionally modeled using all-pole or pole-zero models.[19] This approach is motivated by the speech production model. In contrast, in this paper, we model the phase and envelope of a bandpass filtered speech signal, over a T second duration, directly in the time domain using poles and zeros. That is, in our case the poles and zeros are located in the complex-time plane, called the ζ plane. If a complex signal has all the zeros inside (outside) the unit circle ($|\zeta|=1$) in the ζ plane, it is called a minimum phase or MinP (maximum phase or MaxP) signal. If the signal has poles and zeros in reciprocal complex conjugate pairs, then the signal is called an all-

phase or AllP signal (similar to an all-pass filter). These types of signal models are the duals of well-known filter types in the systems theory literature. The basic notation for these types of signal models is developed in Sec. II.

(ii) *Zero crossings associated with certain analytic signals:* In Sec. III we show that the real (or imaginary) part of MinP or MaxP or AllP signals are RZ signals. That is, if the zero crossings of these RZ signals are known, then the corresponding MinP or MaxP or AllP signals can be reconstructed from the zero-crossing locations, to within a scale factor and a frequency translation. For a reader who is familiar with speech analysis literature, we point out that these zero crossings are the time-domain analogs of what is known as sine spectral frequencies (LSF) in linear prediction analysis.[20,21]

(iii) *Decomposition of arbitrary analytic bandpass signals into component analytic signals:* In Sec. IV we show that an arbitrary bandpass signal can be decomposed into MinP/MaxP and AllP signals by a model fitting method that is analogous to the well-known all-pole or LPC method in speech analysis. An important distinction is that in our case the all-pole modeling is accomplished in the ζ plane instead of the traditional complex z plane. We call this approach inverse signal analysis.[18] This result sets the stage for representing arbitrary bandpass signals by CoZeCs and hence extends Logan's work.

(iv) *Zero-crossing representation algorithm for real-valued bandpass signals:* In Sec. V, we apply the results obtained in Secs. III and IV to real-valued signals. The key result in this section is that if a real-valued bandpass signal has negligible energy in the low-frequency region of the spectrum then the MinP and MaxP parts of the underlying analytic signal can be represented by CoZeCs without actually computing the corresponding analytic signal. A computer simulation of an algorithm that extracts these zero crossings, called the RZC algorithm, is given to illustrate the basic idea.

(v) *Filter banks for speech signal representation:* In Sec. VI we have applied the RZC algorithm to speech signals. It is shown that the speech signal can be reconstructed given the CoZeCs. Conclusions are presented in Sec. VII.

II. DUALITY BETWEEN SIGNALS AND SYSTEMS

In this section we propose rational signal models to describe the envelope and phase of an analytic signal.[22] In traditional engineering literature, linear time-invariant, continuous-time systems are described by a rational system function,

$$H(s)=c_0\prod_{k=1}^{Q}(s-z_k)\Big/\prod_{k=1}^{P}(s-p_k),$$

where s is the complex-frequency variable, defined as $s\triangleq\sigma+j\omega$, $j=\sqrt{-1}$. p_k and z_k are the poles and zeros of the system. From the pole/zero plot one could often get a sense of the frequency response of the system, $H(j\omega)$, immediately. Analogously, for discrete-time systems, a system function $H(z)$ is defined as

$$H(z)=c_1\prod_{k=1}^{Q}(z-z_k)\Big/\prod_{k=1}^{P}(z-p_k),$$

where z is the corresponding complex-frequency variable. In this case the frequency response of the system is the function $H(z)$ evaluated around the unit circle $|z|=1$.[23] The frequency response of the discrete time system is 2π periodic. In the above cases the frequency is regarded as a complex variable. Analogously, we could also regard *time as a complex variable* and thereby define a complex-time (τ) plane, where $\tau\triangleq\sigma+jt$. In the τ plane, we may model a nonperiodic complex-valued signal as

$$x(\tau)=c_2\frac{\Pi_{k=1}^{Q}(\tau-z_k)}{\Pi_{k=1}^{P}(\tau-p_k)} \tag{1}$$

given sufficient number of poles p_k and zeros z_k.

Analogous to the frequency response $H(j\omega)$, the signal $x(t)$ is obtained by evaluating $x(\tau)$ along the jt axis. Carrying the above analogy further, the dual of a complex-frequency z plane, is a complex-time ζ plane, suitable for modeling complex-valued periodic signals. In this case the signal function in terms of poles and zeros is

$$x(\zeta)=c_3\frac{\Pi_{k=1}^{Q}(\zeta-z_k)}{\Pi_{k=1}^{P}(\zeta-p_k)}. \tag{2}$$

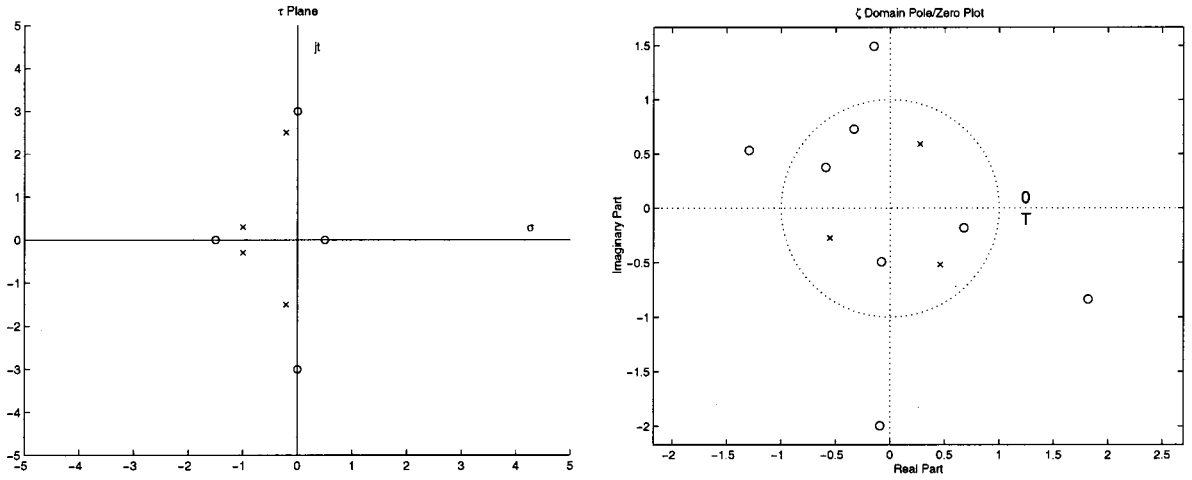
We obtain the periodic signal $x(t)$ by evaluating $x(\zeta)$ around the unit circle $|\zeta|=1$, i.e., $\zeta=e^{-j\Omega t}$, where $\Omega=2\pi/T$ is the fundamental frequency and T is the period. Hence, the unit circle in the ζ plane corresponds to the time interval 0 to T seconds. Figure 2 shows typical pole/zero plots in the two complex-time planes. From the location of poles and zeros in the ζ plane, we can generally infer where in time (0 to T seconds) the peaks and troughs in the envelope of $x(t)$ are located. Voelcker[24] called this way of modeling signals as a ''product representation of signals.'' Also refer to recent work by Poletti,[25] Picinbono,[26] and Kumaresan.[18]

Further, the concept of causality in the systems domain (i.e., the impulse response of a causal system is zero for negative time) is the dual of analyticity in the signal domain [i.e., the spectrum of an analytic signal $x(t)$ is zero for negative frequency]. Also, the group delay (the negative of the derivative of the phase response of a system with respect to frequency) is the dual of instantaneous frequency (IF) (the time derivative of the phase) of $x(t)$. In the next section we shall consider periodic and analytic signal models that are analogs of finite impulse response (FIR) systems. Real-valued signals are dealt with in a later section.

A. FIR-like signal models in the ζ plane

Consider a periodic analytic signal $s_a(t)$, with period T seconds. Let $\Omega=2\pi/T$ denote its fundamental angular frequency. If $s_a(t)$ has a finite bandwidth, it may be described by the following model for a sufficiently large M, over an interval of T seconds:

$$s_a(t)=e^{j\omega_l t}\sum_{k=0}^{M}a_k e^{jk\Omega t}, \tag{3}$$



(a) (b)

FIG. 2. Poles and zeros in complex-time planes: The τ plane is suitable for modeling nonperiodic signals and the ζ plane for modeling periodic signals.

where $\omega_l \geqslant 0$, which represents a frequency translation, is the low-frequency band edge, that we take to be an integer multiple of Ω, say $\omega_l = K\Omega$. a_k are the complex amplitudes of the sinusoids $e^{jk\Omega t}$; $a_0, a_M \neq 0$. By analytic continuation we may regard $s_a(t)$ as the function $s_a(\zeta) = \zeta^{-K}(a_0 + a_1\zeta^{-1} + a_2\zeta^{-2} + \cdots a_M\zeta^{-M})$ evaluated around the unit circle, $\zeta = e^{-j\Omega t}$. In $s_a(\zeta)$ we use the negative powers of ζ in order to maintain the analogy with the traditional use of negative powers of z in the complex-frequency domain. We may factor this polynomial into its $M(=P+Q)$ factors and rewrite $s_a(t)$ as

$$s_a(t) = a_0 e^{j\omega_l t} \underbrace{\prod_{i=1}^{P} (1 - p_i e^{j\Omega t})}_{\text{MinP}} \underbrace{\prod_{i=1}^{Q} (1 - q_i e^{j\Omega t})}_{\text{MaxP}}; \tag{4}$$

where $p_1, p_2, \ldots, p_P$, and $q_1, q_2, \ldots, q_Q$ denote the polynomial's roots; $p_i = |p_i| e^{j\theta_i}$, $q_i = |q_i| e^{j\phi_i}$ and $|p_i| < 1$ and $|q_i| > 1$. Thus, p_i's denote roots inside the unit circle in the complex plane, and q_i are outside the unit circle. Currently we assume that there are no roots on the unit circle. Each factor of the form $(1 - p_i e^{j\Omega t})$ in the above is called an "elementary signal." [24] The p_i and q_i are referred to as (nontrivial) zeros of the signal $s_a(t)$. The above expressions, representing a bandlimited periodic signal, are, of course, the counterpart of the frequency responses of the standard finite impulse response (FIR) filters.[27]

The factors corresponding to the zeros inside the unit circle, $\Pi_{i=1}^{P}(1 - p_i e^{j\Omega t})$, constitute the minimum phase (MinP) signal. Similarly the factors corresponding to the zeros outside the circle, $e^{j\omega_l t}\Pi_{i=1}^{Q}(1 - q_i e^{j\Omega t})$, constitute the (frequency translated) maximum phase (MaxP) signal. These are the direct counterparts of the frequency responses of the well-known minimum and maximum phase FIR filters in discrete-time systems theory.[23] Just as in systems theory (see Sec. 10.3 in Ref. 23) the phase of the MinP signal is the Hilbert transform of its log envelope. That is, the MinP signal may be expressed in the form $e^{\alpha(t) + j\hat{\alpha}(t)}$, where $\hat{\alpha}(t)$ is the Hilbert transform of $\alpha(t)$. See Ref. 18 for details. Similarly, since a maximum phase (MaxP) signal has zeros outside the unit circle, it may be expressed as $e^{\beta(t) - j\hat{\beta}(t)}$, where $\hat{\beta}(t)$ is the Hilbert transform of $\beta(t)$. Thus, the envelope or phase alone is sufficient to essentially characterize a MinP or a MaxP signal. Along the same lines, an all-phase (AllP) analytic signal (the analog of an all-pass filter) would be of the form $e^{j\psi(t)}$. Thus, $s_a(t)$ may be expressed as

$$s_a(t) = A_c e^{j(\omega_l + Q\Omega)t} \underbrace{e^{\alpha(t) + j\hat{\alpha}(t)}}_{\text{MinP}} \underbrace{e^{\beta(t) - j\hat{\beta}(t)}}_{\text{MaxP}}. \tag{5}$$

A_c is $a_0 \Pi_{i=1}^{Q}(-q_i)$. The formulas for $\alpha(t)$ and $\beta(t)$ depend on the particular values of p_i and q_i, respectively. See Ref. 18 for details.

Just as the MixP systems (with zeros inside and outside the unit circle) may be decomposed into all-pass and minimum phase systems (see Sec. 5.6 in Ref. 23), so too $s_a(t)$ may be decomposed into two component signals. Note that in Eq. (4) the zeros, q_i and p_i are assumed to be outside and inside the unit circle, respectively. To obtain the AllP factor, we shall reflect the q_i to inside the circle (as $1/q_i^*$) and cancel them using poles. Then we may group all the zeros inside the unit circle to form a different MinP signal and the zeros outside the circle and the poles that are their reflections inside the unit circle to form the all-phase or AllP part of the signal. That is,

$$s_a(t) = a_0 e^{j\omega_l t} \underbrace{\prod_{i=1}^{P} (1 - p_i e^{j\Omega t}) \prod_{i=1}^{Q} \left(1 - \frac{1}{q_i^*} e^{j\Omega t}\right)}_{\text{MinP}} \times \underbrace{\frac{\Pi_{i=1}^{Q}(1 - q_i e^{j\Omega t})}{\Pi_{i=1}^{Q}\left(1 - \frac{1}{q_i^*} e^{j\Omega t}\right)}}_{\text{AllP}}. \tag{6}$$

Equivalently, multiplying and dividing Eq. (5) by $e^{j2\hat{\beta}(t)}$ and collecting terms, we get

$$s_a(t)=A_c e^{j\omega_l t}\underbrace{e^{\alpha(t)+\beta(t)+j[\hat{\alpha}(t)+\hat{\beta}(t)]}}_{\text{MinP}}\underbrace{e^{j[\Omega_0 t+\phi_1-2\hat{\beta}(t)]}}_{\text{AllP}}. \tag{7}$$

This grouping of signals is, of course, analogous to the well-known decomposition of a linear discrete-time system into minimum phase and all-pass systems (Sec. 5.6 in Ref. 23). Analogous to the fact that the group delay of the all-pass filters is always positive (Sec. 5.5 in Ref. 23), the instantaneous frequency (IF) of the AllP part will always be positive[18] (even if ω_l, the lower band edge, is zero). Hence we[18] called the IF of the AllP part the positive instantaneous frequency or PIF. Later in this paper we use the above models to represent the envelopes and phases of successive overlapping segments of a signal. A real bandpass signal $s(t)$ is modeled as the real part of $s_a(t)$. For a slowly varying signal one can imagine that the p_i and q_i are slowly drifting parameters that characterize the signal's envelope and phase variations. We wish to capture in certain zero-crossing locations the behavior of the slowly varying parameters p_i and q_i.

III. ZERO CROSSINGS THAT CHARACTERIZE ANALYTIC BANDPASS SIGNALS

In this section we show that the real (or imaginary) part of analytic bandpass signals (i.e., the MinP, MaxP, and AllP signals) we introduced in the previous section are RZ signals, i.e., their zero crossings are sufficient to reconstruct these signals.

A. Zero crossings related to minimum/maximum phase signals

Consider a MinP signal, $h_m(t)$, defined as follows:

$$h_m(t)=h_0+h_1e^{j\Omega t}+h_2e^{j2\Omega t}+\cdots+h_me^{jm\Omega t}. \tag{8}$$

An analytic continuation of $h_m(t)$ in the ζ plane is denoted by $H_m(\zeta)$,

$$H_m(\zeta)=h_0+h_1\zeta^{-1}+h_2\zeta^{-2}+\cdots+h_m\zeta^{-m}. \tag{9}$$

Since $h_m(t)$ is MinP, the roots of $H_m(\zeta)$ lie strictly inside the unit circle. Let $H_m^*(1/\zeta^*)$ denote the reciprocal polynomial (with roots in reciprocal conjugate locations, i.e., outside the unit circle):

$$H_m^*(1/\zeta^*)\triangleq h_0^*+h_1^*\zeta^1+h_2^*\zeta^2+\cdots+h_m^*\zeta^m. \tag{10}$$

We define two other polynomials using $H_m(\zeta)$ and $H_m^*(1/\zeta^*)$:

$$P(\zeta)=\zeta^{p/2}H_m(\zeta)+\zeta^{-p/2}H_m^*(1/\zeta^*), \tag{11}$$

$$Q(\zeta)=\zeta^{p/2}H_m(\zeta)-\zeta^{-p/2}H_m^*(1/\zeta^*). \tag{12}$$

Note that the coefficients of $P(\zeta)$ and $Q(\zeta)$ have conjugate-even and conjugate-odd symmetry, respectively. We now show that if $p\geqslant m$, all the roots of $P(\zeta)$ and $Q(\zeta)$ are on the unit circle and interlaced with each other. This result is a direct analog of results known in the speech literature as "line spectrum frequencies."[20,21]

Rewriting Eqs. (11) and (12) in a product form, we have

$$P(\zeta)=\zeta^{p/2}H_m(\zeta)[1+G(\zeta)], \tag{13}$$

$$Q(\zeta)=\zeta^{p/2}H_m(\zeta)[1-G(\zeta)], \tag{14}$$

wherein $G(\zeta)$ is an all-pass or all-phase function,

$$G(\zeta)\triangleq\zeta^{-p}\frac{H_m^*(1/\zeta^*)}{H_m(\zeta)}. \tag{15}$$

$G(\zeta)$ can be factored as

$$G(\zeta)=e^{j(\Omega\tau_0+m\pi)}\zeta^{-(p-m)}\prod_{i=1}^{m}\frac{\zeta_i^*-\zeta^{-1}}{1-\zeta_i\zeta^{-1}}, \tag{16}$$

where ζ_i's are the roots of $H_m(\zeta)$. $\zeta_i=r_ie^{j\Omega\tau_i}$, and $r_i<1$. $\Omega\tau_0=\angle(h_0^*/h_0)$. Since $G(\zeta)$ is an all-pass function, we can write

$$G(\zeta)|_{\zeta=e^{-j\Omega t}}=e^{j\psi(t)}. \tag{17}$$

It should be clear from Eqs. (13) and (14) that $P(\zeta)$ and $Q(\zeta)$ have roots at the locations where $e^{j\psi(t)}$ equals -1 and 1, respectively.

The phase function $\psi(t)$ can further be expressed as follows (similar to the phase of all-pass filters as in Sec. 5.5 in Ref. 23):

$$\psi(t)=\Omega\tau_0+m\pi+(p-m)\Omega t+\sum_{i=1}^{m}2\tan^{-1}\left(\frac{r_i\sin[\Omega(t+\tau_i)]}{1-r_i\cos[\Omega(t+\tau_i)]}\right). \tag{18}$$

The instantaneous frequency, $f(t)$, of $G(e^{-j\Omega t})$ is $(1/2\pi)\times[d\psi(t)/dt]$ and is given by

$$f(t)=\frac{\omega}{2\pi}\left((p-m)+\sum_{i=1}^{m}\frac{1-r_i^2}{|1-r_ie^{j[\Omega(t+\tau_i)]}|^2}\right). \tag{19}$$

If $p\geqslant m$, and since all $r_i<1$, we conclude that $f(t)>0$, i.e., $f(t)$ is a PIF. Therefore $\psi(t)$ is a monotonically increasing function. Let ϕ_0 denote the phase of $G(e^{-j\Omega t})$ at $t=0$, i.e., $\psi(0)=\phi_0$, and $\psi(2\pi/\Omega)=\phi_0+2p\pi$. Therefore, $\psi(t)$ crosses lines corresponding to each integer multiple of π [odd and even multiples of π for $P(\zeta)$ and $Q(\zeta)$, respectively] exactly once, resulting in $2p$ crossing points for $0\leqslant\Omega t<2\pi$. Because the solution to $P(\zeta)=0$ or $Q(\zeta)=0$ requires that $G(\zeta)=\pm1$, these points constitute the total $2p$ roots of $P(\zeta)$ and $Q(\zeta)$ alternately on the unit circle.[20,21]

Since $H_m(\zeta)$ is MinP, the phase of $H_m(\zeta)$ (when evaluated around the unit circle $|\zeta|=1$) and its log envelope are related by the Hilbert transform.[18,23] That is,

$$H_m(\zeta)|_{\zeta=e^{-j\Omega t}}=e^{\gamma(t)+j\hat{\gamma}(t)}, \tag{20}$$

where the phase function $\hat{\gamma}(t)$ is the Hilbert transform of the log-magnitude function $\gamma(t)$. Similarly, evaluating $H_m^*(1/\zeta^*)$ around the unit circle we have

$$H_m^*(1/\zeta^*)|_{\zeta=e^{-j\Omega t}}=e^{\gamma(t)-j\hat{\gamma}(t)}. \tag{21}$$

Plugging Eq. (20), Eq. (21) and $\zeta=e^{-j\Omega t}$ in Eq. (11) and Eq. (12), we have

$$p(t)=P(e^{j\Omega t})=2e^{\gamma(t)}\cos\left(\frac{p}{2}\Omega t-\hat{\gamma}(t)\right). \tag{22}$$

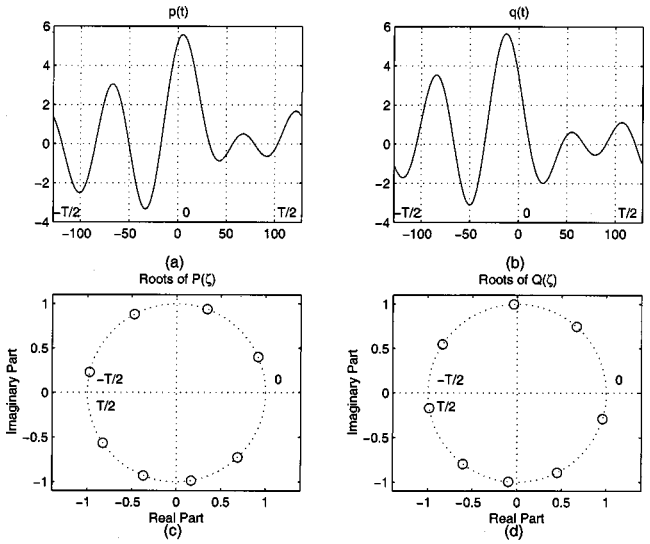


FIG. 3. Real-zero (RZ) signals related to MinP signals: The $p(t)$ and $q(t)$ calculated from a minimum phase signal $h_m(t)$ are plotted in (a) and (b). The roots/zeros of corresponding $P(\zeta)$ and $Q(\zeta)$ with $p=8>m$ are displayed in (c) and (d).

Similarly,

$$q(t)=jQ(e^{j\Omega t})=2e^{\gamma(t)}\sin\left(\frac{p}{2}\Omega t-\hat{\gamma}(t)\right). \tag{23}$$

Since $e^{\gamma(t)}$ has no real zero crossings, all real zero crossings of $p(t)$ and $q(t)$ are due to the cosine and the sine term.

Given the zero-crossing locations $t_1,t_2,\ldots,t_p$ corresponding to $p(t)$, we can compute the roots $e^{j\Omega t_1},e^{j\Omega t_2},\ldots,e^{j\Omega t_p}$. Then the product of the factors $(1-e^{j\Omega t_i}\zeta^{-1})$, $i=1,2,\ldots,p$ gives $P(\zeta)$ (up to a scale factor). Similarly, one obtains $Q(\zeta)$ from the zero crossings of $q(t)$. Using $P(\zeta)$ and $Q(\zeta)$ [in Eqs. (11) and (12)], we can determine $H_m(\zeta)$ and hence $h_m(t)$. *Thus, only zero-crossing information is sufficient to reconstruct signals that are the real (or imaginary) part of frequency translated MinP signals. As mentioned before, such signals are called real zero (RZ) signals*. If the given $h_m(t)$ is a MaxP signal, we can simply interchange the roles of $H_m(\zeta)$ and $H_m^*(1/\zeta^*)$ in the previous discussion, and all the above results are still valid.

A simple example is shown in Fig. 3. We picked a MinP signal $h_m(t)$ with

$$\begin{aligned}H_m(\zeta)=&1.0+(0.6931-j1.7071)\zeta^{-1}+(-1.2025\\&-j0.7020)\zeta^{-2}+(-0.2317+j0.4913)\zeta^{-3}\\&+(0.1436+j0.0461)\zeta^{-4}+(0.0002\\&-j0.0290)\zeta^{-5}+(-0.0056-j0.0003)\zeta^{-6}\\&+(0.0002+j0.0007)\zeta^{-7},\end{aligned}$$

where $m=7$. $P(\zeta)$ and $Q(\zeta)$ were calculated from Eq. (11) and Eq. (12), where $p=8$. Their corresponding RZ signals, $p(t)$ and $q(t)$, are plotted in Figs. 3(a) and 3(b). The roots/zeros of $P(\zeta)$ and $Q(\zeta)$ are shown in Figs. 3(c) and 3(d), respectively. Note that $P(\zeta)$ and $Q(\zeta)$ have all their zeros on the unit circle and they are interlaced. Note also the relationship between the roots of $P(\zeta)$ and $Q(\zeta)$ in the ζ plane with the zero crossings of $p(t)$ and $q(t)$.

B. Zero crossings related to all-phase signals

All-phase (AllP) signals are analytic signals that have both poles and zeros. If the poles are inside the unit circle [as in $E(e^{-j\Omega t})$ below], then the spectrum of the signal is confined to the positive side of the frequency axis (analogous to causal IIR filters). If the poles are outside the unit circle [as in $F(e^{-j\Omega t})$ below], then the spectrum of the signal is confined to the negative side of the frequency axis. Consider an all-phase (AllP) signal $E(e^{-j\Omega t})$ defined as follows:

$$E(e^{-j\Omega t})=\frac{1}{|\Pi_{i=1}^{Q}(-q_i)|}\frac{\Pi_{i=1}^{Q}(1-q_ie^{j\Omega t})}{\prod_{i=1}^{Q}\left(1-\frac{1}{q_i^*}e^{j\Omega t}\right)}. \tag{24}$$

As before, by analytic continuation we can write $E(\zeta)$ as follows:

$$E(\zeta)=e^{j\phi_1}\zeta^{-Q}\frac{B(\zeta)}{B^*(1/\zeta^*)}, \tag{25}$$

where $B(\zeta)\triangleq\Pi_{i=1}^{Q}(1-q_i^{-1}\zeta)$. One may verify Eq. (24) by substituting $\zeta=e^{-j\Omega t}$ in Eq. (25). The roots of $B(\zeta)$ are q_i, $i=1,2,\ldots,Q$, with $1/q_i^*=r_ie^{j\Omega\tau_i}$, and $r_i<1$. Since all the roots of $B(\zeta)$ fall outside the unit circle, $B(\zeta)$ is a MaxP signal and the $B^*(1/\zeta^*)$ is a MinP signal. Clearly, $E(\zeta)$ is already in the form of $G(\zeta)$ encountered in the previous section. Hence, the instantaneous frequency of $E(e^{-j\Omega t})$ is always positive and the phase function of $E(e^{-j\Omega t})$ is a monotonically increasing function. Therefore the zeros of $1+E(\zeta)$ and $1-E(\zeta)$ have properties identical to those of polynomials $P(\zeta)$ and $Q(\zeta)$ discussed in the previous section. That is, $1+E(\zeta)$ and $1-E(\zeta)$ each have Q zeros on the unit circle and they are interlaced. Further, using the notation in Eqs. (6) and (7) we can write $E(e^{-j\Omega t})$ as follows:

$$E(e^{-j\Omega t})=e^{j[Q\Omega t+\phi_1-2\hat{\beta}(t)]}. \tag{26}$$

Thus the unit magnitude root locations of $1\pm E(\zeta)$ correspond to zero crossings of the waveform $1\pm\cos[Q\Omega t+\phi_1-2\hat{\beta}(t)]$ or the waveform $\sin[Q\Omega t+\phi_1-2\hat{\beta}(t)]$. We shall define the imaginary part of $E(e^{-j\Omega t})$ as $e_3(t)$:

$$e_3(t)=\sin[Q\Omega t+\phi_1-2\hat{\beta}(t)]. \tag{27}$$

Given the zero-crossing locations $t_1,t_2,\ldots,t_{2Q}$ corresponding to $\sin[Q\Omega t+\phi_1-2\hat{\beta}(t)]$ or $1\pm\cos[Q\Omega t+\phi_1-2\hat{\beta}(t)]$, we can compute the roots $e^{j\Omega t_1},e^{j\Omega t_2},\ldots,e^{j\Omega t_{2Q}}$. Then we shall define $P_B(\zeta)=\Pi_i(1-e^{j\Omega t_i}\zeta^{-1})$, where the set $\{i\}$ consists of odd integers $1,3,5,\ldots,2Q-1$. Similarly, we shall define $Q_B(\zeta)=\Pi_i(1-e^{j\Omega t_i}\zeta^{-1})$, where the set $\{i\}$ consists of even integers $2,4,6,\ldots,2Q$. Using $P_B(\zeta)$ and $Q_B(\zeta)$ [similar to Eqs. (11) and (12)], we can determine $B(\zeta)$ and hence $E(\zeta)$ to within a scale factor. *Thus, the zero-crossing information of $e_3(t)$ alone is sufficient to reconstruct the AllP signal $E(e^{-j\Omega t})$ up to a complex scale factor*. Hence $e_3(t)$ is a RZ signal.

Similarly, we may consider

$$F(e^{-j\Omega t})=\left|\prod_{i=1}^{P}(-p_i^*)\right|\frac{\prod_{i=1}^{P}(1-p_i e^{j\Omega t})}{\prod_{i=1}^{P}\left(1-\frac{1}{p_i^*}e^{j\Omega t}\right)}. \tag{28}$$

The zeros of $F(e^{-j\Omega t})$ are p_i, $i=1,2,...,P$. $p_i=g_i e^{j\Omega\tau_i}$, and $g_i<1$. The poles of $F(e^{-j\Omega t})$ are outside the unit circle. Thus $F(e^{-j\Omega t})$ has a spectrum confined to the negative side of the frequency axis and its IF is always negative (NIF). As before, $1+F(e^{-j\Omega t})$ and $1-F(e^{-j\Omega t})$ each have P zeros on the unit circle that are interlaced. Again, if the zero crossings of $e_4(t)$,

$$e_4(t)=\sin[P\Omega t+\phi_2-2\hat{\alpha}(t)], \tag{29}$$

are known, then we can reconstruct (using the same algorithm described above) the AllP signal $F(e^{-j\Omega t})$ to within a scale factor. We will make use of $e_3(t)$ and $e_4(t)$ in Sec. V B.

In summary, in this section we have shown that the zero crossings of certain special functions implicitly represent the underlying analytic signals. In other words, the real or imaginary parts of the MinP, MaxP, and AllP signals, are RZ signals, since they are essentially characterized by their real zero crossings. In general, analytic signals are neither MinP/MaxP nor AllP, but mixed-phase (MixP) signals. Hence we have to first decompose an arbitrary MixP signal into MinP/MaxP and AllP signals as shown in Eqs. (6) and (7). An elegant algorithm for achieving this decomposition is presented next. In Sec. V we consider real-valued signals.

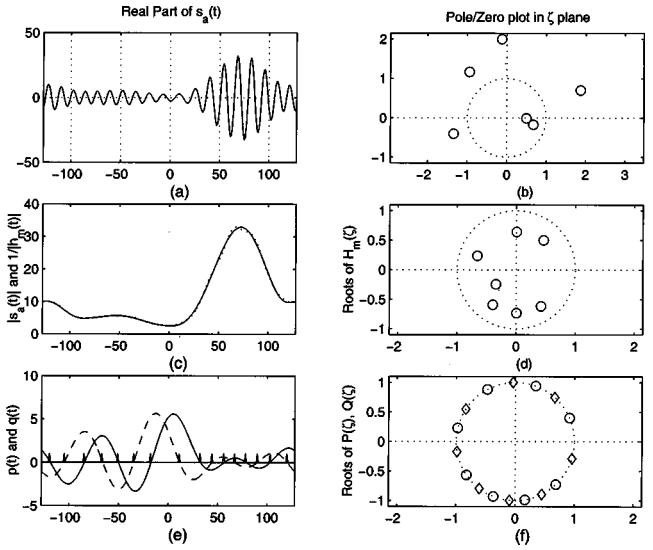


FIG. 4. Envelope of a MixP signal represented by zero crossings: A complex signal $s_a(t)$ is synthesized with six zeros (four outside the unit circle and two inside the unit circle). The zeros of $s_a(t)$ are plotted in (b). The real part of $s_a(t)$ is plotted in (a). The roots of $H_m(\zeta)$ calculated using the LPSD algorithm are shown in (d). Note that $H_m(\zeta)$ is MinP. The estimated envelope $1/|h_m(t)|$ is shown (solid line) in (c); the true envelope is shown by a dotted line. In (e), both RZ functions $p(t)$ and $q(t)$ are plotted by a solid line and a dashed line, respectively. Since they are described fully (to within a scale factor) by their zero crossings, we can represent $p(t)$ and $q(t)$ by only marking their zero-crossing time locations. We show those locations by spikes along the time axis in (e). The roots of $P(\zeta)$ and $Q(\zeta)$, all on the unit circle, are displayed in (f). The roots of $P(\zeta)$ are denoted by a "○" and those of $Q(\zeta)$ are denoted by a "◇."

IV. SEPARATING THE MINP AND ALLP PARTS OF AN ANALYTIC BANDPASS SIGNAL USING LPSD

In this section we present a simple algorithm called the linear prediction in the spectral domain (LPSD).[18] The details of the LPSD algorithm, which separates the MinP and AllP components of a MixP signal, were presented in Ref. 18. Here we summarize these results for completeness. Consider the MixP signal in Eq. (3) or Eq. (4):

$$s_a(t)=e^{j\omega_l t}\sum_{k=0}^{M}a_k e^{jk\Omega t} \tag{30}$$

$$=a_0 e^{j\omega_l t}\prod_{i=1}^{P}(1-p_i e^{j\Omega t})\prod_{i=1}^{Q}(1-q_i e^{j\Omega t}). \tag{31}$$

Using the notation in Sec. II, we may express $s_a(t)$ as

$$s_a(t)=|A_c|e^{[\alpha(t)+\beta(t)]}e^{j((\omega_l+Q\Omega)t+\hat{\alpha}(t)-\hat{\beta}(t)+\phi_1)}. \tag{32}$$

Note that $A_c e^{[\alpha(t)+\beta(t)]}$ is the envelope of $s_a(t)$. The LPSD algorithm separates the MinP and AllP components of MixP signal. This decomposition is achieved by minimizing the energy in an error signal $e(t)$ that is defined as $e(t)=h_m(t)s_a(t)$. The energy in $e(t)$ is defined as follows:

$$\int_0^T|e(t)|^2\,dt=\int_0^T|s_a(t)h_m(t)|^2\,dt. \tag{33}$$

$h_m(t)$ is synthesized using the following formula:

$$h_m(t)=h_0+h_1e^{j\Omega t}+h_2e^{j2\Omega t}+\cdots+h_m e^{jm\Omega t}, \tag{34}$$

where $\Omega=2\pi/T$. Note that $h_m(t)$ is identical to that defined in Sec. III A. The LPSD algorithm minimizes the energy in the error signal $e(t)$ [the integral in Eq. (33) is replaced by a discrete approximation] by choosing the coefficients h_l, where h_0 is constrained to be 1. Thus, the above minimization problem is the direct analog of the autocorrelation method of the linear prediction well-known in spectral/speech analysis[28] as LPC or all-pole modeling or inverse filtering. $h_m(t)$ is the analog of the inverse filter used in LPC and hence is called the "inverse signal."[18] The LPSD algorithm finds an inverse signal $h_m(t)$ such that the envelope of the error signal $e(t)$ is flattened. This can be achieved if the order m of $h_m(t)$ is sufficiently high. m has to be large if there are deep nulls in the signal envelope. After the minimization, since the error signal $e(t)$ has a constant envelope,

$$h_m(t)\approx e^{-[\alpha(t)+\beta(t)]}e^{-j[\hat{\alpha}(t)+\hat{\beta}(t)]}. \tag{35}$$

Figure 4 gives an illustrative example. An analytic signal $s_a(t)$ was synthesized from Eq. (3) using seven Fourier coefficients ($M=6$): $a_0=1$, $a_1=-0.6024-3.2827i$, $a_2=-5.6441+1.5835i$, $a_3=-0.1454+7.4390i$, $a_4=6.4822-1.1832i$, $a_5=-4.6306-6.7388i$, $a_6=1.0737+2.7369i$. The real part of $s_a(t)$ is plotted in Fig. 4(a). The analytic signal $s_a(t)$ has two zeros inside the unit circle and four outside, as shown in Fig. 4(b). Note that $h_m(t)$ computed using the LPSD algorithm is always a MinP signal for any order m, i.e., all of its zeros are inside the unit circle. Figure 4(d) displays the roots of $H_m(\zeta)$. This result is well known in spectral analysis literature.[28] The estimated envelope

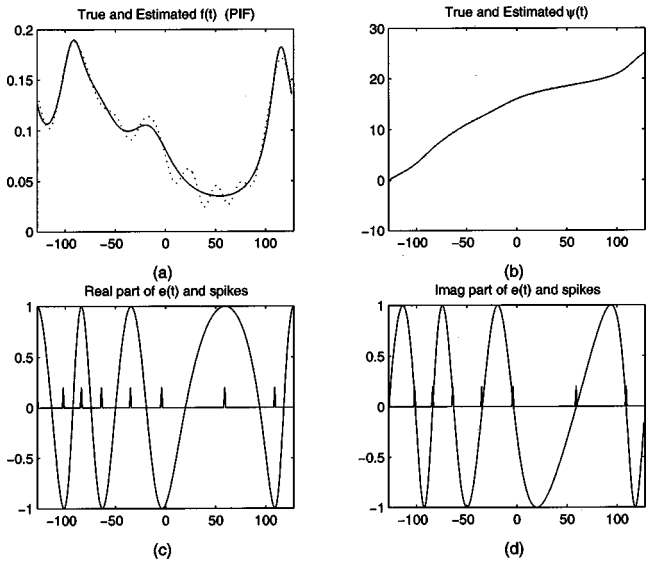


FIG. 5. All-phase (AllP) signal represented by zero crossings: The IFs (i.e., PIFs) of $e(t)$ [both true (solid line) and estimated] are shown in (a); the phase of $e(t)$, denoted by $\psi(t)$, is plotted in (b); because the instantaneous frequency of $e(t)$ is positive, the phase $\psi(t)$ (with $\omega_l=0$) is a monotonically increasing function. The real part and imaginary part of $e(t)$ are shown in (c) and (d), respectively. The indicated spike locations are sufficient information for reconstructing the AllP signal except for a scale factor and a frequency translation. Note that in (c) the spikes correspond to the locations when the real part of $e(t)$ equals ± 1.

$1/|h_m(t)|$ is shown (solid line) in Fig. 4(c); the true envelope, $|s_a(t)|$, is also shown using a dotted line. In Fig. 4(e), both RZ functions $p(t)$ and $q(t)$ [computed using Eqs. (11), (12), (22), and (23) with $p=8>m$] are plotted with a solid line and a dashed line, respectively. Since they are described fully (to within a scale factor) by their zero crossings, we can represent $p(t)$ and $q(t)$ by only marking their zero-crossing time locations. We show these locations in Fig. 4(e), by a train of "spikes" along the time axis. Note that we can uniquely reconstruct $h_m(t)$ from these spike locations. Comparing the envelope in Fig. 4(c) and the spike train in Fig. 4(e), note that when the envelope is large, the density of spikes [due to both $p(t)$ and $q(t)$ together] around that time location is higher. The zeros of $P(\zeta)$ (denoted by a "○") and $Q(\zeta)$ (denoted by a "◇") are displayed in Fig. 4(f). From the above we conclude that the envelope part of the analytic signal $s_a(t)$ has been successfully converted into two RZ signals or two spike trains.

The error signal $e(t)=s_a(t)h_m(t)$ obtains the approximation to the AllP part in Eq. (7),

$$e(t)\approx e^{j[(\omega_l+Q\Omega)t+\phi_1-2\hat{\beta}(t)]}=e^{j\psi(t)}. \tag{36}$$

Note that $e(t)$ is identical to the function $E(e^{-j\Omega t})$ in Sec. III B, except for the frequency translation term $e^{j\omega_l t}$. $e(t)$ (with $\omega_l=0$) and its real and imaginary parts are shown in Fig. 5. The PIF of $e(t)$ is shown in Fig. 5(a). The phase of $e(t)$, denoted by $\psi(t)$, is plotted in Fig. 5(b). Because the IF of $e(t)$ is always positive, the phase $\psi(t)$ is a monotonically increasing function. The real part of $e(t)$ (i.e., $\cos[(Q\Omega)t+\phi_1-2\hat{\beta}(t)]$) and its imaginary part, $\sin[(Q\Omega)t+\phi_1-2\hat{\beta}(t)]$ are shown in Fig. 5(c) and Fig. 5(d), respectively. As explained in Sec. III B [between Eqs. (26) and (27)] the alternate zero crossings $\sin[(Q\Omega)t+\phi_1-2\hat{\beta}(t)]$ occur at the locations where $\cos[(Q\Omega)t+\phi_1-2\hat{\beta}(t)]$ reaches 1 and -1. These locations are indicated by spikes in Fig. 5(c) and Fig. 5(d). There is sufficient information to reconstruct $e(t)$ to within a scale factor and a frequency translation. Figure 6 summarizes the steps involved in representing a mixed-phase (MixP) analytic signal by zero crossings.

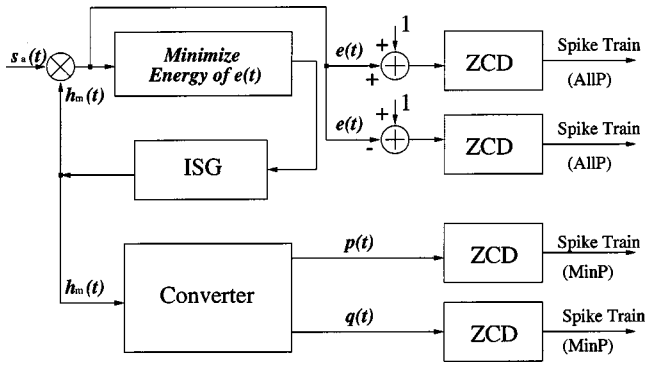


FIG. 6. Algorithm for representing an analytic signal by zero-crossings: The algorithm for representing an analytic signal by zero crossings of RZ signals associated with the MinP/MaxP and AllP parts. ISG stands for "inverse signal generator," ZCD stands for the "zero-crossing detector." "Converter" block computes $p(t)$ and $q(t)$ from $h_m(t)$ using Eqs. (11) and (12).

In summary, in this section we have shown that we can use the LPSD algorithm to separate an analytic signal into MinP and AllP components, each of which can be represented by CoZeCs, as described in Sec. III. However, the above results are only applicable to analytic signals. In practice, one observes real-valued signals. Is it possible to represent real-valued bandpass signals directly by CoZeCs without having to compute their analytic version by using a Hilbert transformer? The answer is yes and in the next section we describe the algorithm.

V. EXTENDING THE ZERO-CROSSING REPRESENTATION TO REAL-VALUED BANDPASS SIGNALS

In this section we assume that we are given a real-valued bandpass signal $s(t)$. Let $s(t)$ be defined as the real part of $s_a(t)$ given in Eq. (3):

$$s(t)=\sum_{k=0}^{M}|a_k|\cos[(\omega_l+k\Omega)t+\theta_k]. \tag{37}$$

(Recall that $\omega_l=K\Omega$, is the lower band edge.) Our main purpose in this section is to extend the results in Secs. III and IV (which were meant for analytic signals) to processing and representing real-valued bandpass signals by modifying the inverse signal approach slightly, taking into account the fact that real-valued signals have spectral components on both sides of the frequency axis. In Sec. V A we show that the LPSD algorithm can be directly applied to $s(t)$ provided that the low-frequency region of the spectrum of $s(t)$ has negligible energy. This sets the stage for Sec. III B, which describes the RZC algorithm. In Sec. V B we define two real-valued inverse signals, $q(t)$ and $r(t)$, such that the error signals $e_1(t)=s(t)q(t)$ and $e_2(t)=s(t)r(t)$ when low-pass filtered, result in RZ signals $e_3(t)$ and $e_4(t)$. The zero-

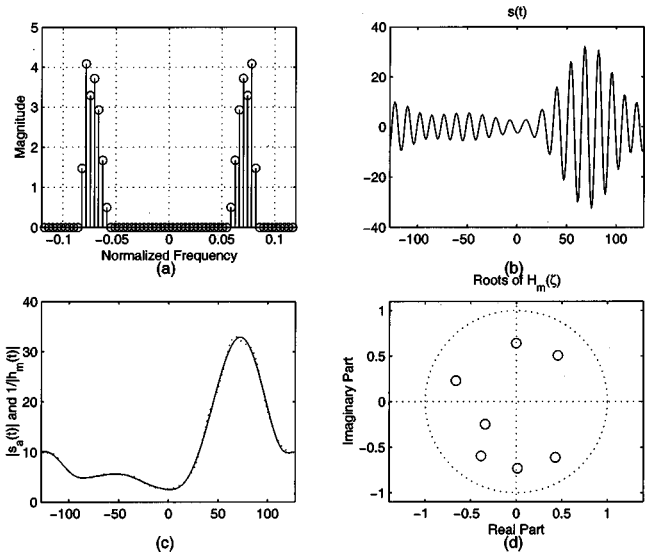


FIG. 7. Inverse signal that represents the envelope of a real signal: (a) shows the magnitude spectrum of original real signal $s(t)$ [the real part of the analytic signal $s_a(t)$ used in Fig. 4]; the time waveform of $s(t)$ is plotted in (b); (c) shows the estimated envelope together with the true envelope $|s_a(t)|$. The roots of $H_m(\zeta)$ are plotted in (d). $m=7$.

crossing locations of $e_3(t)$ and $e_4(t)$ are sufficient to reconstruct the MinP and MaxP parts of $s_a(t)$ and hence characterize $s(t)$.

A. Computing the inverse signal $h_m(t)$ from a real bandpass signal $s(t)$

Consider $s(t)$ over an interval of 0 to T seconds. We shall rewrite $s(t)$ for convenience as follows:

$$s(t)=\sum_{k=-N}^{N} b_k e^{jk\Omega t}. \tag{38}$$

Since $s(t)$ is real-valued, $b_{-k}=b_k^*$. Comparing Eqs. (37) and (38), we note that $b_{K+i}=a_i$, for $i=0,1,...,M$. $N=K+M$. $\omega_h=N\Omega$ is the higher band edge. Let $s(t)$, be such that the Fourier coefficients $b_{-K+1},...,b_0,...,b_{K-1}$ are equal to zero for some $K<N$. An example of the spectrum is shown in Fig. 7(a). Following the development in Sec. IV, let us define an error signal $e(t)$ over 0 to T seconds as follows:

$$e(t)=s(t)h_m(t), \tag{39}$$

where $h_m(t)=1+\Sigma_{l=1}^{m} h_l e^{jl\Omega t}$. As before, our goal is to find an inverse signal, $h_m(t)$ (i.e., choose the coefficients, h_l), such that the energy in the error signal $e(t)$ is minimized. Plugging in the expression for $s(t)$ from Eq. (38) into the error-energy expression, we get

$$\int_0^T |e(t)|^2\,dt=\int_0^2 |s(t)h_m(t)|^2\,dt \tag{40}$$

$$=T\sum_{n=-N}^{N+m} |g_n|^2, \tag{41}$$

where $g_n=b_n*h_n$ ($*$ denotes linear convolution), and $h_0=1$. The inverse signal coefficients, h_l, can be determined by solving linear equations using the LPSD algorithm.

Since the Fourier coefficients b_{-K+1} to b_{K-1} are all assumed to be zero, the expression for the error energy can be written as a sum of two terms, if $m\leqslant 2K-1$:

$$\int_0^T |e(t)|^2\,dt=T\sum_{n=-N}^{-K+m} |g_n|^2+T\sum_{n=K}^{N+m} |g_n|^2 \tag{42}$$

$$=\int_0^T |[s(t)-j\hat{s}(t)]h_m(t)|^2\,dt$$

$$+\int_0^T |[s(t)+j\hat{s}(t)]h_m(t)|^2\,dt. \tag{43}$$

Since the analytic $[s_a(t)=s(t)+j\hat{s}(t)]$ and antianalytic $[s_a^*(t)=s(t)-j\hat{s}(t)]$ signals are complex conjugates of each other, the two terms in the above expression are equal. Thus the inverse signal obtained by minimizing any one of the terms in the above expression is equal to the $h_m(t)$ obtained by minimizing the error in Eq. (40) [using the real-valued $s(t)$]. Note that the second integral in Eq. (42) is identical to the one in Eq. (33). This is a crucial observation, because the $h_m(t)$ obtained by minimizing the error energy in Eq. (40) is not only a MinP signal, but further, $1/|h_m(t)|$ also gives the Hilbert envelope of the analytic signal $s_a(t)$, although it is computed directly from the real-valued $s(t)$. However, for the latter to be true, $s(t)$ must have sufficient number of zero-valued Fourier coefficients in the low-frequency region, i.e., m must be less than or equal to $2K-1$.

Some simulation results are provided in Fig. 7. A real signal $s(t)$ [the real part of the analytic signal $s_a(t)$ used in Fig. 4] is plotted in Fig. 7(b). It was synthesized using seven Fourier coefficients, whose magnitudes are shown in (a). In this simulation, we let $K=15$ to ensure that there are sufficient number of zero-valued Fourier coefficients in the low-frequency region. The envelope of the signal was estimated using the above algorithm for $m=7<2K-1$. The resulting $h_m(t)$ is MinP, so all the roots are inside the unit circle, as shown in Fig. 7(d). Figure 7(c) displays the envelope estimate against the true envelope obtained from the analytic signal. The higher the value of m (as long as $m\leqslant 2K-1$) the better is the approximation. The above method tends to match the peaks of the envelope much more precisely than the valleys. This behavior is well known in traditional LP analysis.[19]

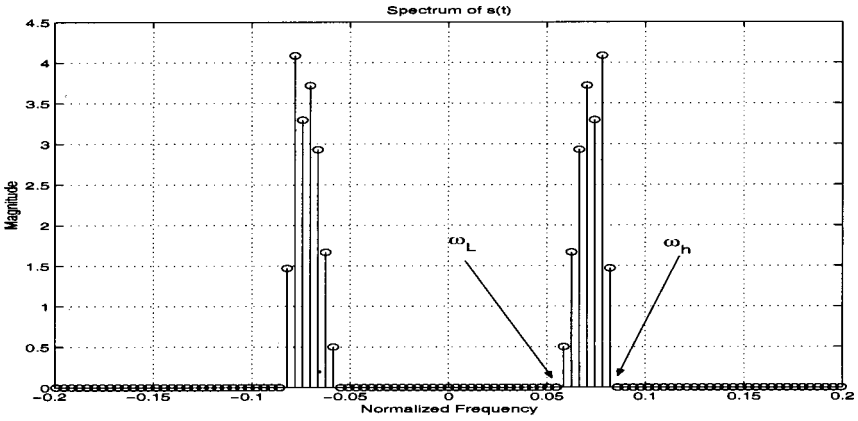
B. Computation of CoZeCs that represent the bandpass signal $s(t)$

In this section we show that if we choose two real-valued inverse signals $q(t)$ and $r(t)$ (defined below) appropriately, then we can obtain two RZ signals whose zero crossings in effect represent $s(t)$ up to a scale factor and a frequency translation. Again, consider $s(t)$ over an interval of 0 to T seconds. Now let us define $q(t)$ and $r(t)$,

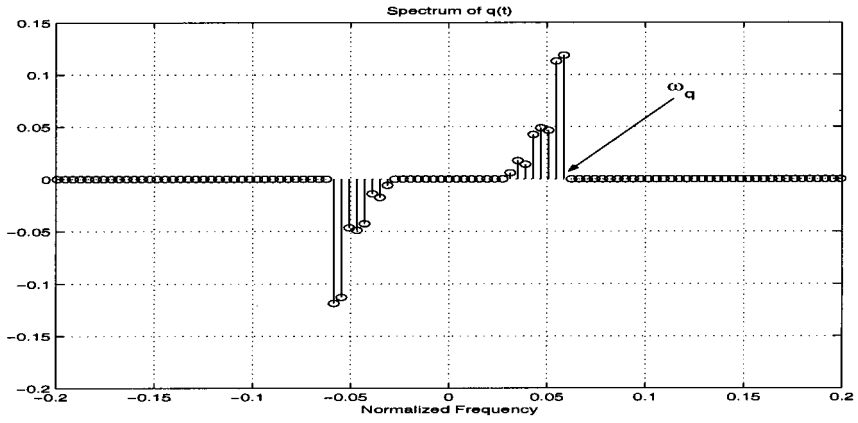
$$q(t)=[e^{j\omega_q t}h_m(t)-e^{j\omega_q t}h_m^*(t)]/2j, \tag{44}$$

$$r(t)=[e^{j\omega_r t}h_m(t)-e^{-j\omega_r t}h_m^*(t)]/2j. \tag{45}$$

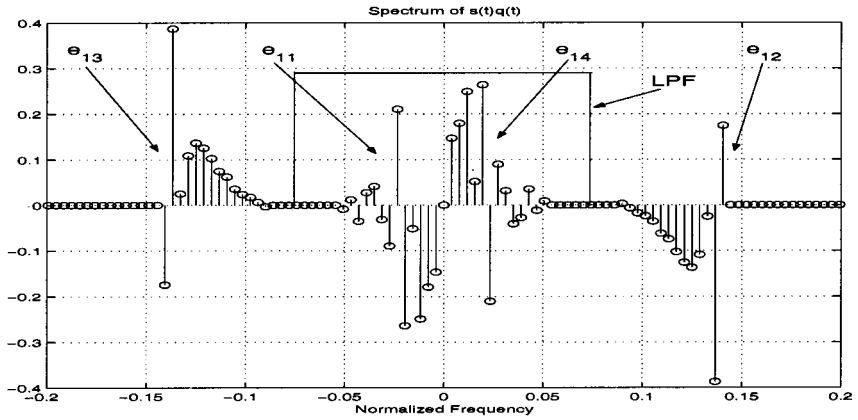
ω_r and ω_q are specified later, but they are integer multiples of Ω. Further, let us define two error signals, $e_1(t)$ and $e_2(t)$:



(a)



(b)



(c)

FIG. 8. The spectrum associated with $e_1(t)$: (a) The magnitude spectrum of real signal $s(t)$; (b) the magnitude spectrum of real signal $q(t)$; (c) the magnitude spectrum of error signal $e_1(t)=s(t)q(t)$.

$$e_1(t)=s(t)q(t), \tag{46}$$

$$e_2(t)=s(t)r(t). \tag{47}$$

Our goal is to compute $q(t)$ and $r(t)$ such that the energy in the error signals $e_1(t)$ and $e_2(t)$ is minimized. In other words, we shall minimize $\int_0^T|e_1(t)|^2\,dt$ [and/or $\int_0^T|e_2(t)|^2\,dt$] by choosing the coefficients, h_l. Since these two error energies are almost the same, here we take $e_1(t)$ as an example. The error-energy expression for $e_1(t)$ is

$$\int_0^T|e_1(t)|^2\,dt=\int_0^T|s(t)q(t)|^2\,dt. \tag{48}$$

The inverse signal coefficients, h_l, can be determined by solving linear equations similar to those in LPSD. Refer to the example in Fig. 8. Let us rewrite $s(t)$ and $q(t)$ in terms of their analytic and antianalytic parts as $s(t)=1/2[s(t)+j\hat{s}(t)+s(t)-j\hat{s}(t)]$ and $q(t)=(1/2j)\{\hat{q}(t)+jq(t)-[\hat{q}(t)-jq(t)]\}$. Then the error can be expressed as the sum of four terms as follows, provided the Fourier coeffi-

cients corresponding to each of the four kernels do not overlap:

$$\int_0^T |e_1(t)|^2 dt = \frac{1}{16}\int_0^T |(s(t)-j\hat{s}(t))(\hat{q}(t)+jq(t))|^2 dt + \frac{1}{16}\int_0^T |(s(t)-j\hat{s}(t))(\hat{q}(t)-jq(t))|^2 dt$$

$$+\frac{1}{16}\int_0^T |(s(t)+j\hat{s}(t))(\hat{q}(t)+jq(t))|^2 dt + \frac{1}{16}\int_0^T |(s(t)+j\hat{s}(t))(\hat{q}(t)-jq(t))|^2 d \tag{49}$$

$$= \frac{1}{16}\int_0^T |\underbrace{(s(t)-j\hat{s}(t))(e^{j\omega_q t}h_m^*(t))}_{e_{11}}|^2 dt + \frac{1}{16}\int_0^T |\underbrace{(s(t)-j\hat{s}(t))(e^{-j\omega_q t}h_m(t))}_{e_{12}}|^2 dt$$

$$+\frac{1}{16}\int_0^T |\underbrace{(s(t)+j\hat{s}(t))(e^{j\omega_q t}h_m^*(t))}_{e_{13}}|^2 dt + \frac{1}{16}\int_0^T |\underbrace{(s(t)+j\hat{s}(t))(e^{-j\omega_q t}h_m(t))}_{e_{14}}|^2 dt. \tag{50}$$

The spectrum associated with each of the kernels e_{11}, e_{12}, e_{13}, and e_{14} is clearly marked in Fig. 8. [Here we have used the same real signal $s(t)$ shown in Fig. 7.] If this nonoverlap condition is met, then, as in the previous section all the four terms in the above expression will be equal. In that case the inverse signal, $h_m(t)$, obtained by minimizing any one of the terms in the above expression is equal to the $q(t)$ obtained by minimizing the error in Eq. (48) [using the real-valued $s(t)$ and $q(t)$]. This guarantees a MinP $h_m(t)$ and hence $1/|\hat{q}(t)+jq(t)|$ gives an estimate of the Hilbert envelope of the analytic (and the antianalytic) signal. An example is given in Fig. 8. The spectrum of $s(t)$ and $q(t)$ are shown in Figs. 8(a) and (b), respectively. The spectrum of $s(t)q(t)$ is given in Fig. 8(c).

The nonoverlap condition requires that $q(t)$ must have a suitable carrier frequency ω_q. There are two possible overlaps in the spectrum of $s(t)q(t)$, i.e., between e_{11} and e_{14}, and between e_{14} and e_{12}. To avoid overlap between e_{11} and e_{14}, ω_q should be such that, $\omega_q \leqslant \omega_l$. In order to be able to determine $q(t)$ uniquely from the coefficients h_l and *vice versa* [see Eq. (44)], ω_q should be greater than $(m+1)\Omega$. To avoid overlap between e_{14} and e_{12} (or e_{11} and e_{13}) we should choose m, the order of $q(t)$, such that $m\Omega < \omega_q - 1/2(\omega_h - \omega_l)$. In summary, we should choose $(m+1)\Omega < \omega_q \leqslant \omega_l$ and $m\Omega < \omega_q - 1/2(\omega_h - \omega_l)$. Similar comments also apply to $r(t)$, which is a real-valued signal on the higher-frequency side of $s(t)$. (See Fig. 9.) In this case, $\omega_r \geqslant \omega_h$ to avoid the overlap between e_{21} and e_{24}; and $m\Omega < 2\omega_l$ to avoid the overlap between e_{24} and e_{22}.

The real signal $s(t)$ could also be written in terms of envelope and phase as follows [see Eq. (32)]:

$$s(t) = |A_c| e^{[\alpha(t)+\beta(t)]} \cos[(K\Omega + Q\Omega)t + \hat{\alpha}(t) - \hat{\beta}(t) + \phi_1]. \tag{51}$$

The real signals $q(t)$ and $r(t)$ calculated by the above-mentioned process are

$$q(t) = -\text{imag}\{e^{-[\alpha(t)+\beta(t)]} e^{-j[\omega_q t + \hat{\alpha}(t) + \hat{\beta}(t)]}\} \tag{52}$$

$$= -e^{-[\alpha(t)+\beta(t)]} \sin[K\Omega t + \hat{\alpha}(t) + \hat{\beta}(t)], \tag{53}$$

$$r(t) = -\text{imag}\{e^{-[\alpha(t)+\beta(t)]} e^{j[\omega_r t - \hat{\alpha}(t) - \hat{\beta}(t)]}\} \tag{54}$$

$$= -e^{-[\alpha(t)+\beta(t)]} \sin[N\Omega t - \hat{\alpha}(t) - \hat{\beta}(t)], \tag{55}$$

where we chose $\omega_q = K\Omega$ and $\omega_r = N\Omega$. Then the two error signals are

$$e_1(t) = -|A_c| \sin[(2K\Omega + Q\Omega)t + 2\hat{\alpha}(t) + \phi_1] + |A_c| \sin[Q\Omega t - 2\hat{\beta}(t) + \phi_1], \tag{56}$$

$$e_2(t) = -|A_c| \sin\{[(K+N)\Omega + Q\Omega]t + 2\hat{\beta}(t) + \phi_1\} + |A_c| \sin[P(\Omega t - 2\hat{\alpha}(t) + \phi_2]. \tag{57}$$

Low pass filtering the $e_1(t)$ and $e_2(t)$ with the cut-off frequency $K\Omega$ [refer to Fig. 8(c) and Fig. 9(c)], we have

$$e_3(t) = |A_c| \sin[Q\Omega t - 2\hat{\beta}(t) + \phi_1], \tag{58}$$

$$e_4(t) = |A_c| \sin[P\Omega t - 2\hat{\alpha}(t) + \phi_2]. \tag{59}$$

These two signals are the same as in Eq. (27) and Eq. (29), but for a scale factor. From the discussions in Sec. III B, we know that both $e_3(t)$ and $e_4(t)$ are RZ signals and they determine the corresponding AllP factors. Using these, we can reconstruct the corresponding analytic signals up to a complex scale factor and a frequency translation. The filtered error signals $e_3(t)$ and $e_4(t)$ together with their "true" values are displayed in Figs. 10(a) and (b).

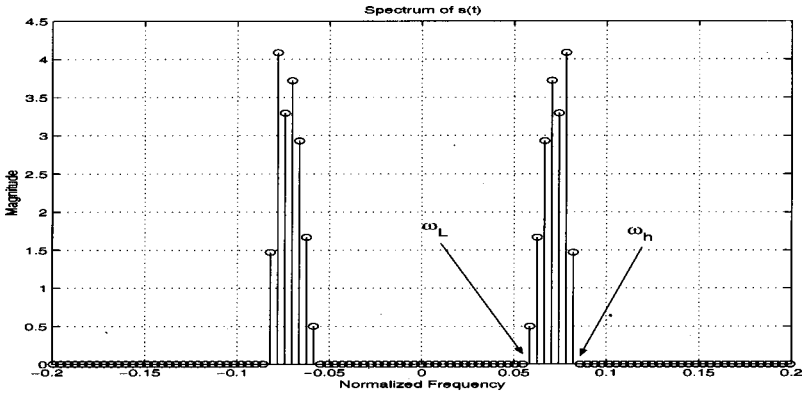
C. Summary of RZC algorithm

The steps involved in the RZC algorithm are listed below and shown in Fig. 11.

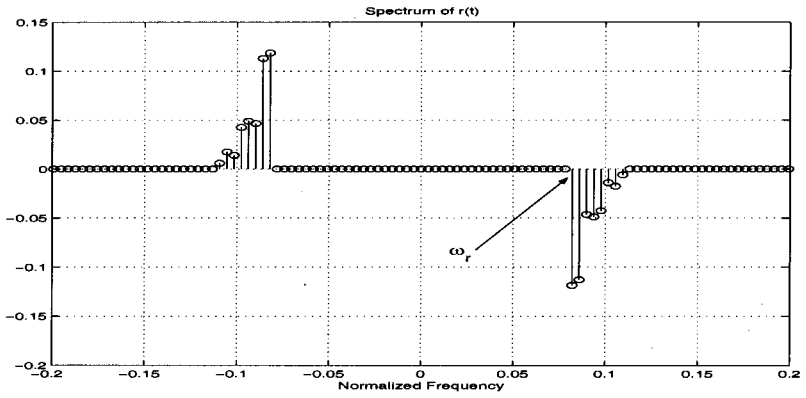
Real Zero Conversion (RZC) Algorithm (Analysis)

Given: Real-valued bandpass signal $s(t)$.

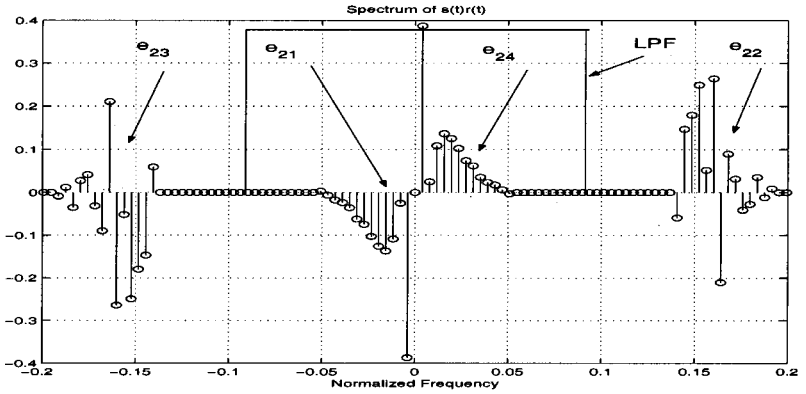
1. Calculate $h_m(t)$ (i.e. the coefficients of $h_m(t)$) by applying LPSD algorithm to real signal $s(t)$
2. Compute $q(t)$ and $r(t)$ from $h_m(t)$ using Eqs. 44 and 45.
3. Compute $e_1(t) = s(t)q(t)$, and $e_2(t) = s(t)r(t)$.
4. Low-pass filter $e_1(t)$ and $e_2(t)$ to produce $e_3(t)$ and $e_4(t)$. Determine the zero-



(a)



(b)



(c)

FIG. 9. The spectrum associated with $e_2(t)$: (a) the magnitude spectrum of real signal $s(t)$; (b) the magnitude spectrum of real signal $r(t)$; (c) the magnitude spectrum of error signal $e_2(t)=s(t)r(t)$.

crossings (CoZeCs) of $e_3(t)$ and $e_4(t)$.
5. Compute the MinP part ($e^{\alpha(t)+j\hat{\alpha}(t)}$) and MaxP part ($e^{\beta(t)-j\hat{\beta}(t)}$) using the CoZeCs (see Section III).
6. Estimate ω_l, $|A_c|$ and ϕ_1 by means of standard least squares using Eqs. 61–63.

Output: Two sets of CoZeCs, ω_l, $|A_c|$ and ϕ_1.

Real Zero Conversion (RZC) Algorithm (Synthesis)

Given: Two sets of CoZeCs, ω_l, $|A_c|$ and ϕ_1.
1. Compute the MinP part ($e^{\alpha(t)+j\hat{\alpha}(t)}$) and MaxP part ($e^{\beta(t)-j\hat{\beta}(t)}$) using the CoZeCs (see Section III). (Same as step 5 in the RZC analysis algorithm).
2. Use the MinP and MaxP parts in conjunction with the estimates of ω_l, $|A_c|$ and ϕ_1, to reconstruct $s(t)$ using Eq. 60.

Output: Estimate of the signal $s(t)$.

Note that the final outcome of the above algorithm is that the $2P$ plus $2Q$ zero-crossing locations corresponding to the filtered error signals $e_3(t)$ and $e_4(t)$ determine the polynomial $s_a(t)$, whose order is $P+Q$. Note that the polynomial $s_a(t)$ is a complex polynomial and hence there are $2P+2Q$ real numbers that determine the polynomial (ignoring

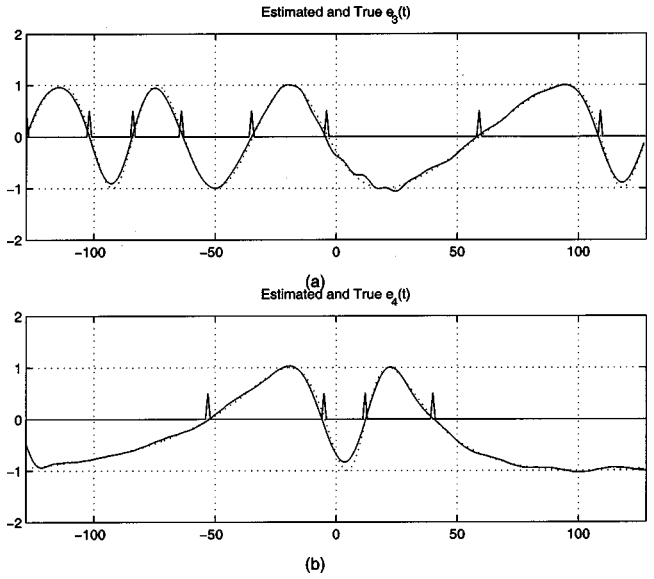


FIG. 10. CoZeCs: (a) Estimated (solid line) and true (dotted line) $e_3(t)$ are plotted. (b) Estimated (solid line) and true (dotted line) $e_4(t)$ are plotted. Note that the number of zero crossings of $e_3(t)$ is $2Q$ and those of $e_4(t)$ is $2P$.

the first coefficient a_0 that is absorbed in the scale factor A_c). Hence the RZC algorithm is a way of transforming the $P+Q$ complex Fourier coefficients corresponding to the trigonometric polynomial that represents $s(t)$ into $2(P+Q)$ zero-crossing locations that implicitly determine the underlying analytic signal $s_a(t)$.

We wish to make clear the conditions under which the above transformation can be achieved. Recall that the key idea in Sec. IV is to flatten the signal envelope by using the all-pole model (LPSD algorithm) thereby turning the error signal $e(t)$ into an AllP signal. The desirable properties associated with the zero crossings ensue from this AllP signal. There are two situations under which it may not be possible to completely flatten the envelope of a bandpass signal. First, if $s(t)$ is such that its envelope dips to zero for some t [i.e., $s_a(t)$ has one or more zeros on the unit circle], then, clearly the LPSD algorithm would require an extremely large m to fit an all-pole model to the signal envelope and hence the nonoverlap conditions mentioned above may not be met. Second, if $s(t)$ is such that the Fourier coefficients in the low-frequency region are not sufficiently small, then again the above nonoverlap conditions are not met. These two conditions may impact the accuracy with which the signal $s(t)$ is reconstructed from the CoZeCs. However, by choosing the filter bank appropriately the situation can be mitigated. It is interesting to note that if m is chosen large, ignoring the above overlap conditions, then $1/|\hat{q}(t)+jq(t)|$, instead of approximating the Hilbert envelope $[|s_a(t)|]$ just approximates the full-wave rectified signal envelope, $|s(t)|$.

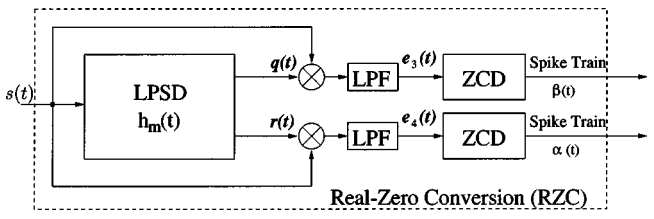


FIG. 11. RZC algorithm for real bandpass signals: The signals $q(t)$ and $r(t)$ are obtained by minimizing the energy in $e_1(t)$ and $e_2(t)$, respectively. In fact, one has to determine $h_m(t)$ and then form $q(t)$ and $r(t)$. The error signals $e_1(t)$ and $e_2(t)$ are low-pass filtered to obtain $e_3(t)$ and $e_4(t)$, which fully represent the original real signal (up to a scale factor and a carrier frequency translation). This block is called the real-zero converter (RZC) in Fig. 1.

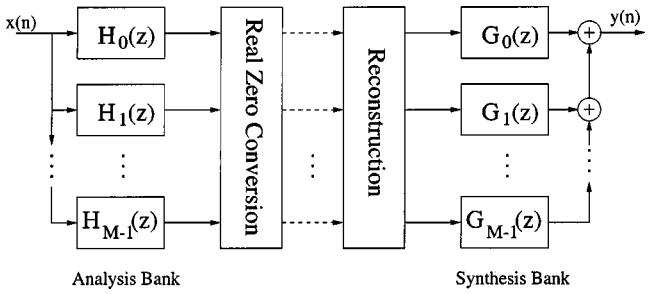


FIG. 12. Analysis–synthesis filter bank: In Sec. VI A the analysis filters are FIR and have the same bandwidth, while in Sec. VI B gamma-tone filters are employed that have different bandwidths.

VI. APPLICATION TO SPEECH ANALYSIS

In order to analyze speech using the RZC algorithm, we first bandpass filter the signal using a filter bank. We split the input speech signal into M uniformly or nonuniformly distributed frequency bands. We use two different filter banks in this section: a linear-phase perfect reconstruction filter bank, and a gamma-tone filter bank that supposedly mimics the cochlear filtering.

A. Perfect reconstruction filter bank

In this section we use a linear phase perfect reconstruction filter bank followed by the RZC algorithm to process a speech signal obtained from the ISOLET database (http://cslu.cse.ogi.edu/corpora/isolet/), *isolet/isolet1/mjc1/mjc1-P2-t.adc*. The speech signal corresponds to the spoken utterance /p/ by a male speaker. The signal waveform [Fig. 15(a)] is sampled at 16 kHz and is 7392 samples long. The signal is first preemphasized using a filter with a transfer function $1-0.98z^{-1}$. Our general speech analysis and synthesis model is shown in Fig. 12. We filter the speech signal using $M=32$ uniformly distributed linear-phase FIR filters. The filters were designed using a Matlab program provided by the Multirate Signal Processing Group at the University of Wisconsin—Madison (http://saigon.ece.wisc.edu/waveweb/QMF.html). The filters $G_k(z)$ all have unity transfer function and $H_k(z)$ are chosen such that $\Sigma_{k=0}^{31}H_k(z)=1$. The order of the filter is 192. The 3-dB bandwidth of each bandpass filter $H_k(z)$ is approximately 250 Hz. The filters closest to dc (1st channel) and the Nyquist frequency (26th to 32nd) are ignored since there is little energy in those regions.

1. Speech analysis

The output of each bandpass filter is viewed through sliding (Hamming) windows shown in Fig. 13. The windows overlap each other by 50%. We set the length of the windows to roughly that of a pitch period. In this example, T

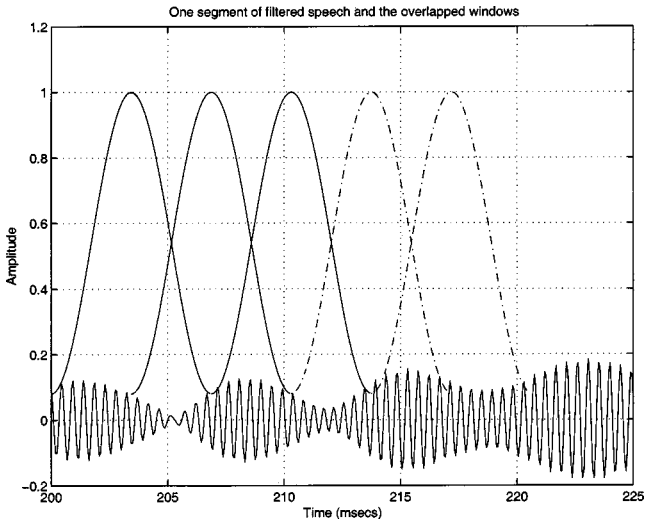


FIG. 13. One segment of filtered speech with overlapped sliding windows: We used a Hamming window (6.9 ms duration). Each window overlaps the previous one by 50%.

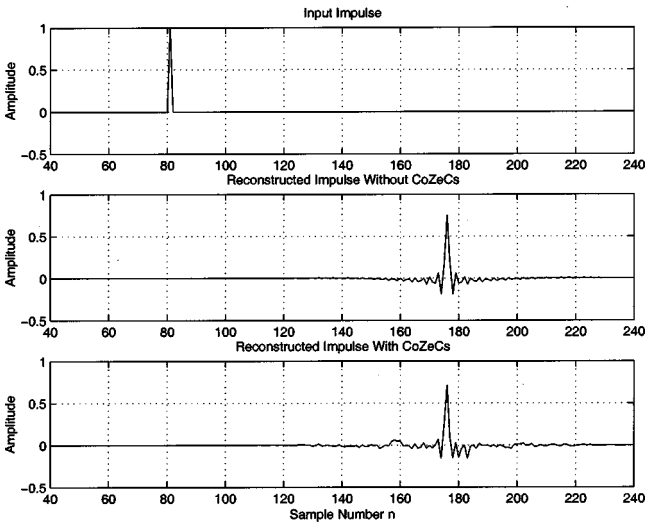


FIG. 14. Reconstruction of an impulse with and without RZC using the linear phase perfect reconstruction filter bank: We compare the reconstructed impulse with and without RZC, where the reconstruction without RZC is plotted in the middle panel and the reconstruction with RZC is displayed in the bottom panel. They are closely matched to each other. The top panel shows the input impulse location. Since the filter bank is a perfect reconstruction variety, we know that if the input is an impulse the output must be an impulse with a delay. The delay is equal to the group delay of the filter (96 samples). However, the output is not exactly an impulse because we omit the filters 1 and 26 to 32 from our filter bank.

$=6.9$ ms, except in the lower-frequency channels (channel 6 and lower), where $T=13.8$ ms ($\approx$ twice the pitch period).

Note that the spectrum of the windowed signal is the convolution of the spectrum of the speech signal and the spectrum of the window function. Hence, using the Hamming window tends to reduce the magnitude of the Fourier coefficients in the low-frequency region as required for the RZC algorithm (see Sec. V B).

We estimate the lower band edge ω_l and the higher band edge ω_h of $s(t)$ from the cut-off frequencies of each bandpass filter. Then we set $\omega_q=\omega_l-\delta\Omega$ and $\omega_r=\omega_h+\delta\Omega$, where $\delta\geqslant 0$. δ is a "guard" band and in this simulation $\delta=1$. The order m of $h_m(t)$ is chosen to be 7 in channels 2 to 6 and 9 in channels 7 to 25. To each block of the windowed signal in each channel, we apply the RZC algorithm. For each block we obtain $e_3(t)$ and $e_4(t)$ as in Eq. (27) and Eq. (29) and determine their zero crossings (CoZeCs). Using these zero crossings we estimate the MinP ($e^{\alpha(t)+j\hat{\alpha}(t)}$) and MaxP ($e^{\beta(t)-j\hat{\beta}(t)}$) components. Note that the model of the real signal is

$$s(t)=|A_c|e^{[\alpha(t)+\beta(t)]}\cos[(\omega_l+Q\Omega)t+\hat{\alpha}(t)-\hat{\beta}(t)+\phi_1]. \tag{60}$$

After we have the estimates of MinP and MaxP components, the estimates of $\hat{\omega}_l$, $|A_c|$, and ϕ_1 are obtained by standard least squares (see, for example, pp. 261–269 in Ref. 29). The least squares estimate of ω_l is obtained by finding the maximum of the function

$$\frac{4}{T}\left|\int_{-T/2}^{T/2}s(t)e^{-[\alpha(t)+\beta(t)]}e^{-j[\omega_l t+Q\Omega t+\hat{\alpha}(t)-\hat{\beta}(t)]}\,dt\right|^2. \tag{61}$$

Magnitude $|A_c|$, and the phase ϕ_1 can be shown to be approximately $A_c=\sqrt{\hat{\alpha}_1^2+\hat{\alpha}_2^2}$ and $\phi_1=\arctan(-\hat{\alpha}_2/\hat{\alpha}_1)$, where

$$\hat{\alpha}_1=\frac{2}{T}\int_{-T/2}^{T/2}s(t)e^{-[\alpha(t)+\beta(t)]}\cos\{-[\hat{\omega}_l t+Q\Omega t+\hat{\alpha}(t)-\hat{\beta}(t)]\}dt, \tag{62}$$

$$\hat{\alpha}_2=\frac{2}{T}\int_{-T/2}^{T/2}s(t)e^{-[\alpha(t)+\beta(t)]}\sin\{-[\hat{\omega}_l t+Q\Omega t+\hat{\alpha}(t)-\hat{\beta}(t)]\}dt. \tag{63}$$

Thus, each channel outputs the zero-crossing locations of $e_3(t)$ and $e_4(t)$ and estimates of A_c, ω_l, and ϕ_1.

2. Synthesis

Using the two sets of zero crossings, we reconstruct α, β, $\hat{\alpha}$, $\hat{\beta}$, and the corresponding MinP and MaxP components. (See the details in Sec. III B.) These are then combined with the estimates ω_l, $|A_c|$, and ϕ_1, to reconstruct the windowed signal as in Eq. (60). Because each window overlaps the other by 50%, adding the reconstructed blocks together gives the reconstructed signal for each channel.

The first example simply uses an impulse input to demonstrate the analysis/synthesis idea. Figure 14 shows the outputs of the filter bank when an impulse is input. In the top panel we just plot the input impulse that is applied at the 81st sample. The output of the filter bank without any RZC processing is shown in the middle panel. Although the filter bank has perfect reconstruction property, the output is not an impulse (but close) because we have included in the filterbank only filters numbered 2 to 25 as mentioned before. The group delay of the filters [the cascade of $H_k(z)$ and $G_k(z)$] is 96 samples. The reconstruction of the impulse when the RZC

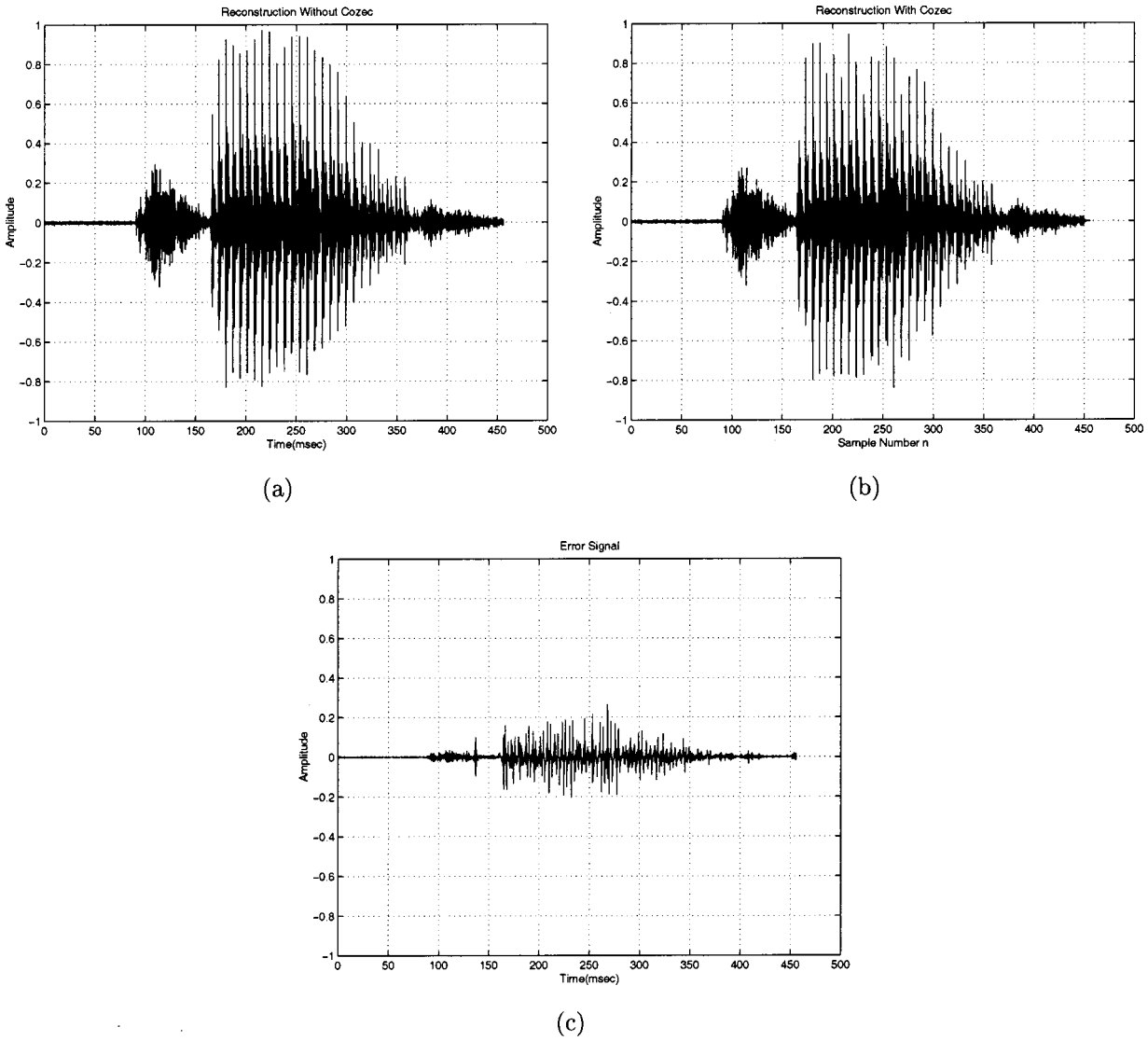


FIG. 15. Reconstruction of the speech signal with and without RZC using a linear phase perfect reconstruction filter bank: We compare the reconstructed speech signal with RZC (b) and without RZC (a). They are closely matched to each other. The difference between these two signals, i.e., the error signal is displayed in (c). The error contribution is larger at lower frequencies.

algorithm is used to represent the filter outputs using CoZeCs is displayed in the bottom panel. It closely matches the one in the middle panel.

We have also compared the reconstructed speech signal with and without RZC in Fig. 15, where the reconstruction without RZC (just add the output from channel 2 to 25 together with a suitable delay) is plotted in the left panel and the reconstruction with the RZC algorithm is displayed in the right panel. They are closely matched to each other. The difference signal obtained by subtracting the reconstructed signal with CoZeCS [Fig. 15(a)] from the reconstructed signal without using CoZeCs [Fig. 15(b)] is plotted in Fig. 15(c). The spectrograms associated with each signal are also displayed in Fig. 16, where (a) is reconstructed without RZC, (b) is reconstructed with RZC and (c) is the spectrogram of the difference signal.

B. Fixed gamma-tone filter bank

In this section we apply the RZC algorithm to a segment of speech (obtained from the TIMIT database) processed through a physiologically motivated auditory filter bank. The speech signal corresponds to the spoken utterance ''She had your dark suit in greasy wash water all year'' by a female speaker (*timit/train/dr1/fcjf0/sa1.wav*). The original waveform is sampled at 16 kHz. We consider only samples with indices 2000 to 9150, which corresponds to the utterance ''She had.'' The signal is preemphasized using a filter with a transfer function $1-0.98z^{-1}$. Our analysis and synthesis system is the same as in Fig. 12, but with different analysis and synthesis filters. The analysis filter bank is similar to those used in many other physiologically motivated auditory models,[30–33] which simulates the motion of the basilar membrane. We use the well-known gamma-tone filter bank,[34] for this purpose. The magnitude responses of 23-channel filters used in this filter bank are shown in Fig. 17(a). It is designed by using an auditory toolbox provided by Malcolm Slaney.[35] As in the previous section the output of each gamma-tone filter is viewed through sliding observation windows. We choose a longer window ($T=27.5$ ms) in the low-frequency band (channels 13–20) and a shorter window ($T=13.8$ ms,

approximately two times the pitch period) in the high-frequency band (channels 4 to 12). The rest of the processing is identical to that described in the previous section.

Given the analysis filters $H_k(z)$, $k=0,...,M-1$, the synthesis filters are chosen as $G_k(z)=H_k^*(z)$, i.e., the synthesis filter bank is just the analysis filter bank with time-reversed impulse response. This leads to good (but not perfect) signal reconstruction. If necessary, an equalizing filter bank can be used to compensate for any imperfections in the reconstruction. Without this equalization we found that the magnitude fluctuations are less than 2 dB [shown in Fig. 17(b)].

We compared the reconstructed signal with and without RZC in Fig. 18, where the reconstruction without RZC is plotted in the top panel and the reconstruction with the RZC algorithm is displayed in the bottom panel. Clearly, the Co-ZeCs represent the speech signal reasonably accurately.

VII. CONCLUSIONS

From Logan's work[11] it is clear that traditional zero crossings of a real-valued bandpass signal cannot uniquely represent it, except in some special cases mentioned in the Introduction. Hence, in this paper, we have sought and found alternate signal-adaptive methods that can be used to re-

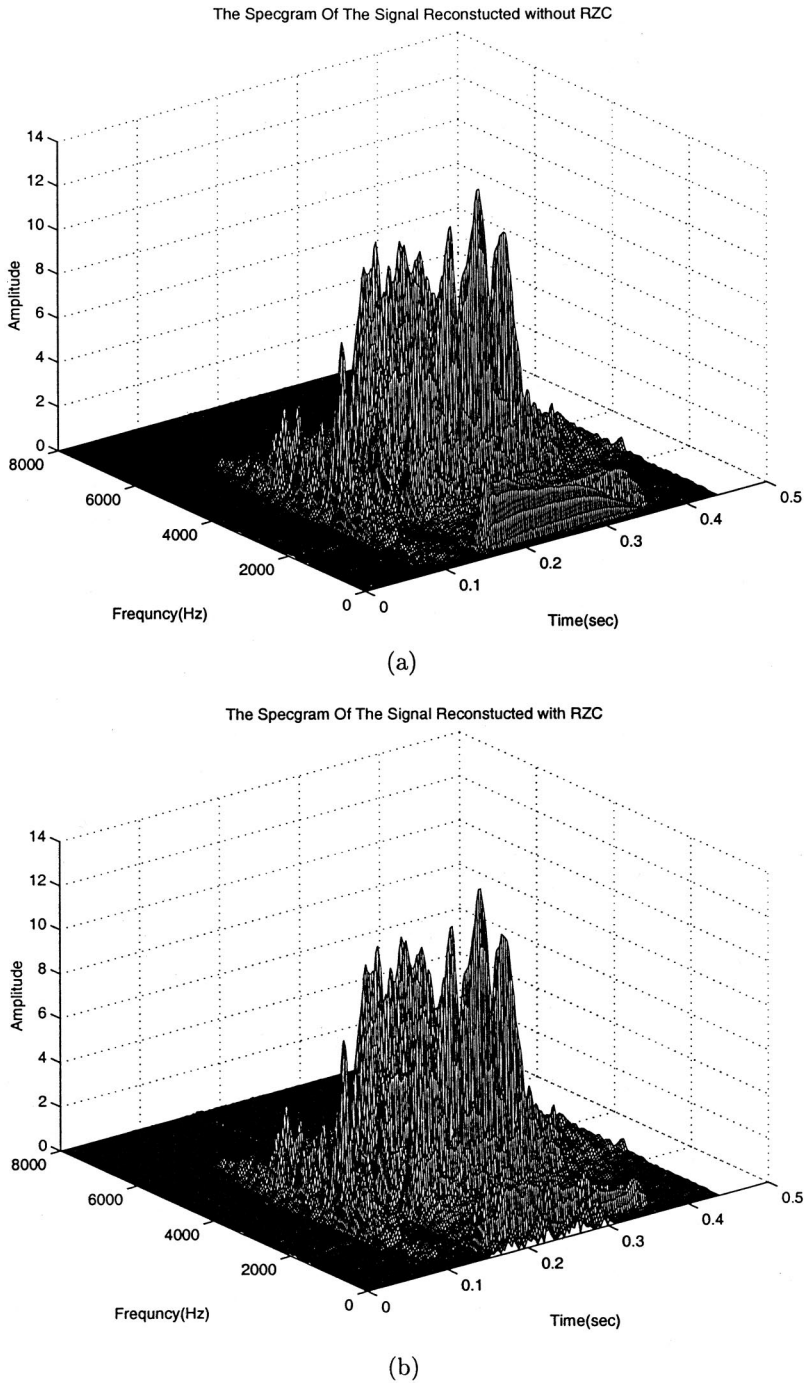


FIG. 16. Spectrograms of the reconstructed speech signal with and without RZC: The figures show the spectrograms of the reconstructed speech signal with and without RZC and the difference signal. Again, note that the error contribution at lower frequencies is larger.

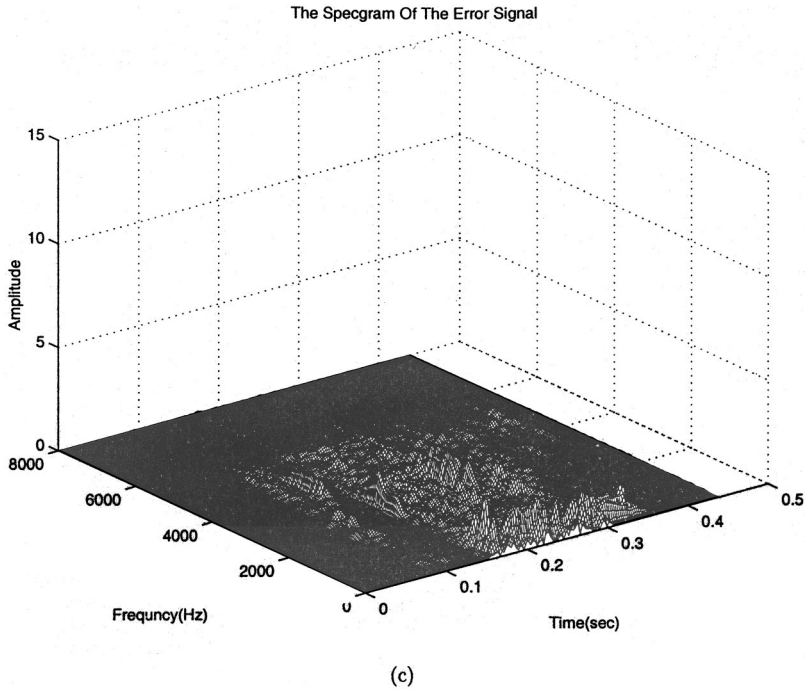


FIG. 16. (*Continued.*)

present arbitrary bandpass signals by timing information (CoZeCs). A key first step is to localize the signal in frequency (by filtering) and in time (by using a T second window) and then compute its representation. The information that needs to be extracted from such a time–frequency localized signal, $s(t)$, consists of A_c, the overall amplitude of the signal, ω_c, the nominal carrier frequency of the signal, and the details of the modulations $e^{\alpha(t)+j\hat{\alpha}(t)}$ and $e^{\beta(t)-j\hat{\beta}(t)}$. Clearly this information is contained in the $P+Q$ complex Fourier coefficients corresponding to the trigonometric polynomial that models $s(t)$ over the time interval T seconds. The proposed RZC algorithm is a way of transforming these $P+Q$ Fourier coefficients into $2(P+Q)$ zero-crossing locations (CoZeCs) that implicitly determine the underlying analytic signal $s_a(t)$ [$s(t)$ is the real part of $s_a(t)$]. A_c and ω_c are constants over a block of T seconds and are expected to vary only slowly from block to block. They constitute the ''place'' information in the parlance of cochlear signal analysis.[5] Typically, A_c and ω_c are the obviously visible information in a spectrogram of the signal. The information about the modulations $e^{\alpha(t)+j\hat{\alpha}(t)}$ and $e^{\beta(t)-j\hat{\beta}(t)}$ (which are not obvious from a spectrogram) are in the zero crossings of $e_3(t)$ and $e_4(t)$ and they constitute the so-called ''rate'' information.[5] In this paper we have used only a fixed filter bank to analyze and represent the signal. Eventually the filters may be made signal adaptive, thereby cutting down the number of filters needed. As mentioned in Sec. III, the line spectrum frequencies (LSFs) used in speech processing to

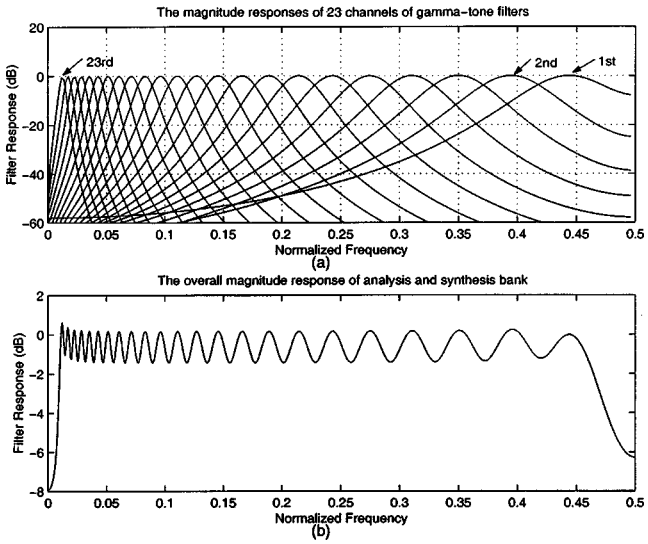


FIG. 17. Magnitude response of the gamma-tone filter bank: Magnitude responses of 23 channels of the gamma-tone filters are show in (a). The overall response of the analysis and synthesis filter bank is shown in (b). The combined analysis–synthesis filter bank exhibits less than 2 dB ripple.

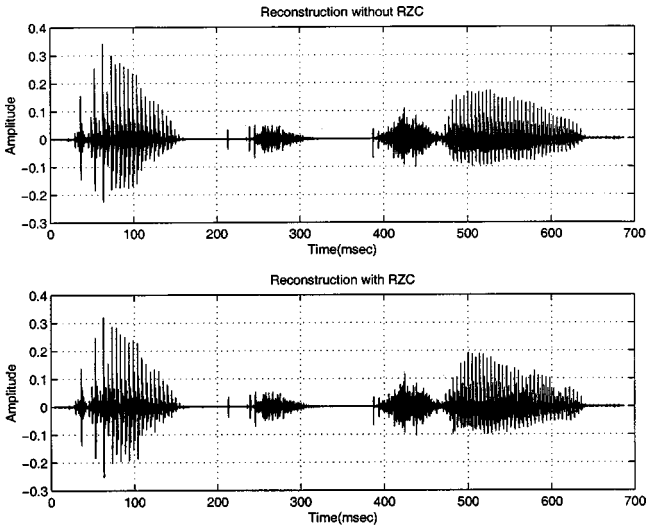


FIG. 18. Reconstruction with and without RZC using the gamma-tone filter bank: Gamma-tone filter bank is used to analyze and synthesize one segment of speech. We compare the reconstructed speech signal without and with RZC, where the reconstruction without RZC is plotted in the top panel and the reconstruction with the RZC is displayed in the bottom panel.

represent the spectral envelope of a signal are the duals of the CoZeCs that represent the signal envelope and phase. It is known that LSFs are statistically robust with respect to quantization and coding,[36] and we expect similar results to hold for CoZeCs as well. This will be a topic for future work.

The proposed approach has a number of potential applications. Although it is not at all certain that the proposed approach has anything to do with the way the inner ear represents signals, the proposed representation is at least an ''existence proof'' that bandpass signals can be represented exclusively by timing information à la the inner ear. Further similarities to the inner-ear signal representation need to be pursued. Second, note that the representation by CoZeCs is an alternative to sub-band coding,[37] in which the signal is processed through a filter bank, down-sampled, and the samples are allotted bits based on the strength of the signal and other perceptual criteria. A similar approach could be applied to the representation by CoZeCs. Third, determining precisely the arrival time differences between spatially distributed receivers is one of the most important issues in estimating the direction of arrival (DOA) of signals. It is easy to visualize a DOA estimator based on comparing the time delays between CoZeCs-based representations of the same signal received at two or more spatially distributed locations.

Finally, it is instructive to compare our approach to the LPC-based methods (such as code-excited linear prediction (CELP)[38]) which are based on a speech production model. CELP coding uses LPC (or equivalently the LSFs) to represent the spectral envelope of the signal. Unfortunately, the spectral phase (or equivalently the prediction error sequence) has contributions from many components (pitch related harmonics) and is hence a ''rough'' function of frequency. Therefore in order to fit the spectral phase, the CELP approach uses an exhaustive search of a code book for a function that is as close to the spectral phase of the prediction-error signal as possible. In contrast, our RZC algorithm captures both the temporal envelope and temporal phase information together in the CoZeCs after decomposing the signal into frequency bands. Further, the pitch information can be extracted by post-processing the CoZeCs from different frequency bands.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Science Foundation under Grant Nos. CCR-9804050 and CCR-0105499. The authors would like to thank Qiang Fu and Jesse Hansen for their help in revising this manuscript. The authors also thank Dr. Shyam Khanna of Columbia university and Dr. Egbert DeBoer of the University of Amsterdam for useful discussions.

[1] A. G. Requicha, ''The zeros of entire functions: theory and engineering applications,'' Proc. IEEE **68**, 308–328 (1980).
[2] A. J. Jerri, ''Shannon sampling theorem—its various extensions and applications: a tutorial review,'' Proc. IEEE **65**, 1565–1596 (1977).
[3] T. V. Sreenivas and R. J. Niederjohn, ''Zero-crossing based spectral analysis and svd spectral analysis for formant frequency estimation in noise,'' IEEE Trans. Signal Process. **40**, 282–293 (1992).
[4] T. Sugihara, S. Kajita, K. Takeda, and F. Itakura, ''On the use of the zero-crossing analysis for multi-channel signal processing,'' J. Acoust. Soc. Am. **100**, 2576 (1996).
[5] H. L. Hawkins *et al.*, in *Auditory Computation* (Springer-Verlag, New York, 1996).
[6] F. Rieke, D. Warland, R. R. van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, MA, 1997).
[7] J. B. Allen, ''Cochlear modeling,'' IEEE Trans. Acoust., Speech, Signal Process. **2**, 3–29 (1985).
[8] L. Abbott and T. J. Sejnowski, *Neural Codes and Distributed Representations: Foundations of Neural Computation* (MIT Press, Cambridge, MA, 1999).
[9] N. E. Hurt, *Phase Retrieval and Zero Crossings* (Kluwer Academic, Norwell, MA, 1989).
[10] H. B. Voelcker and A. Requicha, ''Clipping and signal determinism: Two algorithms requiring validation,'' IEEE Trans. Commun. **COM-21**, 738–744 (1973).
[11] B. F. Logan, ''Information in the zero crossings of band-pass signals,'' Bell Syst. Tech. J. **56**, 487–510 (1977).
[12] S. J. Haavik, ''The conversion of zeros of noise,'' Master's thesis, University of Rochester, Rochester, NY, 1966.
[13] I. Bar-David, ''An implicit sampling theorem for bounded band-limited functions,'' Inf. Control. **24**, 36–44 (1984).
[14] Y. Y. Zeevi, A. Gravriely, and S. Shamai-Shitz, ''Image representation by zero and sine-wave crossings,'' J. Opt. Soc. Am. A **4**, 2045–2060 (1987).
[15] S. Shitz and Y. Y. Zeevi, ''On the duality of time and frequency domain signal reconstruction from partial information,'' IEEE Trans. Acoust., Speech, Signal Process. **33**, 1486–1498 (1985).
[16] F. A. Marvasti, *A Unified Approach to Zero-Crossings and Non-Uniform Sampling* (Nonuniform, Oak Park, IL, 1987).
[17] A. V. Oppenheim, J. S. Lim, and S. R. Curtis, ''Signal synthesis and reconstruction from partial Fourier domain information,'' J. Opt. Soc. Am. **73**, 1413–1420 (1983).
[18] R. Kumaresan and A. Rao, ''Model-based approach to envelope and positive-instantaneous frequency of signals and application to speech,'' J. Acoust. Soc. Am. **105**, 1912–1924 (1999).
[19] J. I. Makhoul, ''Linear prediction: A tutorial review,'' Proc. IEEE **63**, 561–580 (1975).
[20] F. Itakura, ''Line spectrum representation of linear predictive coefficients of speech signal,'' J. Acoust. Soc. Am. Suppl. 1 **57**, S35(A) (1975).
[21] F. K. Soong and B. H. Juang, ''Line spectrum pair (LSP) and speech data compression,'' in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, San Diego, CA, April 1984, pp. 1.10.1–1.10.4.
[22] A. Papoulis, *The Fourier Integral and its Applications* (McGraw-Hill, New York, NY, 1962).
[23] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing* (Prentice–Hall, Englewood Cliffs, NJ, 1989).
[24] H. B. Voelcker, ''Towards a unified theory of modulation part I: Phase-Envelope relationships,'' Proc. IEEE **54**, 340–354 (1966).
[25] M. Poletti, ''The homomorphic analytic signal,'' IEEE Trans. Signal Process. **45**, 1943–1953 (1997).
[26] B. Picinbono, ''On instantaneous amplitude and phase of signals,'' IEEE Trans. Signal Process. **45**, 552–560 (1997).
[27] L. B. Jackson, *Signals, Systems, and Transforms* (Addison-Wesley, Reading, MA, 1991).
[28] S. M. Kay, *Modern Spectral Estimation: Theory and Application* (Prentice–Hall, Englewood Cliffs, NJ, 1987).
[29] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory* (Prentice–Hall, Englewood Cliffs, NJ, 1998).
[30] G. Kubin and W. Kleijn, ''On speech coding in a perceptual domain,'' in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Phoenix, AZ, Mar. 1999, pp. 205–208.
[31] G. Kubin and W. Kleijn, ''Multiple-description coding (mdc) of speech with an invertible auditory model,'' in *IEEE Workshop on Speech Coding*, 1999, pp. 81–83.
[32] S. Seneff, ''A joint synchrony/mean-rate model of auditory speech processing,'' J. Phonetics **16**, 55–76 (1988).
[33] K. W. X. Yang and S. Shamma, ''Auditory representation of acoustic signals,'' IEEE Trans. Inf. Theory **38**, 824–839 (1992).
[34] R. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, and M. Allerhand, ''Complex sounds and auditory images,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Honer (Pergamon, Oxford, 1992), pp. 429–443.

[35] M. Slaney, ''Auditory toolbox,'' Tech. Rep. 45, Apple Computer, Inc., 1994.
[36] J. S. Erkelens and P. M. T. Broerson, ''On the statistical properties of line spectrum pairs,'' in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Detroit, MI, May 1995, pp. 768–771.
[37] R. V. Cox, ''New directions in sub-band coding,'' IEEE Trans. Selected Areas Commun. **6**, 391–409 (1988).
[38] M. R. Schroeder and B. Atal, ''Code-excited linear prediction (celp): High quality speech at very low bit rates,'' in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Tampa, FL, April 1985, p. 937.

Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation

Brian F. G. Katz[a)]
Laboratoire d'Acoustique Musicale, Université de Paris 6, 11 rue de Lourmel, 75015 Paris, France

(Received 21 January 2001; revised 10 July 2001; accepted 20 August 2001)

Human spatial perception of sound is a complex phenomenon. The Head-Related Transfer Function (HRTF) is a vital component to spatial sound perception. In order improve the understanding of the correlation between the HRTF and specific geometry of the head and pinna, a Boundary Element Method (BEM) has been used to calculate a portion of the HRTF of an individual based on precise geometrical data. Advantages of this approach include the ability to alter the geometry of the individual through the model in ways which are not possible with real subjects. Several models are used in the study, including a head with no pinna and several sized spheres. Calculations are performed for various source locations around the head. Results are presented for rigid model cases. Effects of variations on impedance and comparisons to measured data will be presented in the subsequent paper. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1412440]

PACS numbers: 43.64.Bt, 43.66.Qp, 43.66.Pn, 43.64.Ha [LHC]

I. INTRODUCTION

Spatial perception of sound is a very complex phenomenon. It serves in a variety of ways, from localizing a potential attacker to perceiving fullness and envelopment in a concert hall. Determining a mental picture of the surrounding environment without the need to look around, and move around, is an ability which may be unique to humans. One of the key factors in the process of human spatial perception of sound which has come to light over the past few decades is the Head-Related Transfer Function, or HRTF. The head, and especially the pinna (or outer ear), through its complicated shape modifies, attenuates, and amplifies the sound reaching the ear drum in different frequency regions. These changes are dependent on the sound's incident direction. The HRTF relates the measured sound at a position in the auditory canal to the sound that would be measured at the same point in free field (no head) for a given source position.[1] When used correctly, this function is capable of reproducing spatial acoustic scenes over headphones or loudspeakers. It is the first step in understanding how people spatialize and localize, and what further processes are involved. Because the function is based upon the geometrical characteristics of the head and ears, to fully understand the HRTF it is necessary to relate the structural characteristics to the acoustic results. This research hopes to provide another tool for analyzing this relationship. In order to obtain real knowledge of the geometric effects, the ability to change the geometry and monitor the results is almost essential. As it is rather difficult to alter an individual's head and pinna shape, it is proposed that a computer model of the surface of an individual be modeled, and the HRTF be calculated. The computer model has the property that changes can be made to a person's head shape with ease. Such changes could be as drastic as removing the ears entirely. If this model can be verified, using real measurements of the same individual, it could provide a new method for understanding the HRTF, and in so doing result in a better understanding of human spatialization and localization. For practical applications, the HRTF is typically sampled at discrete positions at a fixed distance around the head. In this state, the HRTF becomes a discrete set of filters representing these changes which are described as function of position around an individual.

The work described in this paper entailed the calculation of an individual HRTF from geometrical data. The calculation was based on a boundary element model (BEM) simulation, disregarding any internal structure of the head.[2,3] The geometrical surface model mesh was obtained by using a laser scanning device on an individual subject. This work is part of a larger research project involving the calculation and measurement of individual HRTFs, including the effects of impedance.[3]

II. HISTORICAL WORKS

The amount of research into computational calculations of the sound field around the human head is very limited. A substantial paper was written by Weinrich in 1984.[4] The goal of this work was directed toward a better understanding of the near head sound field for the purpose of hearing aid design and evaluation. The work divided the calculation, and geometry, into several parts. As an introductory experiment in using computational methods for this purpose, many assumptions were made. The ear canal was modeled as a collection of cylinders of varying diameter. The response of the canal was calculated using transmission line theory. A very rough two-dimensional version of the pinna (only the perimeter was modeled using four lines) was created and the response as a function of elevation was calculated using a finite difference time domain approach. The results replicated some of the very basic HRTF effects, primarily the shifting of the primary high frequency notch (around 10 kHz) with varying elevation. The head was modeled using a surface mesh and the field was calculated using a boundary element

[a)]This work was performed while at the Acoustics Department, The Pennsylvania State University, State College, PA 16801. Electronic mail: bfgkatz@attglobal.net

method solution. The response was calculated for the horizontal plane. In designing the mesh, the pinna were not included, resulting in a very coarse mesh with smooth sides. The mesh contained 212 elements, with the maximum spacing between nodes being 5 cm. At four nodes/wavelength, the absolute minimum, this yields a mesh which is usable up to 1.7 kHz. The results, compared to measurements made on a physical replica of the BEM model mesh, were good below this frequency. As the model did not contain the pinna and was very coarse, only very general comparisons were possible to real HRTF measurements. Even so, this work showed that numerical calculations of the human listening response were possible. With the increase in computing power since 1984, the size and complexity of the calculations can be greatly increased, and therefore result in more sophisticated models which may better represent a real individual HRTF.

An additional work in 1988 approximated the head using a sphere or oblate spheroid and calculated the variations between the two using an analytical solution.[5] The result showed little variation between the sphere and spheroids, on the order of 0.5 dB. Comparisons to real HRTFs were not made in this work.

III. BOUNDARY ELEMENT METHOD APPROACH

Calculation of the HRTF is made on the assumption that only the surface characteristics of the head are pertinent; propagation through the head structure is ignored. In determining a method for calculation, several options are available which all vary in complexity and computation requirements. The method chosen here, which is potentially the most efficient manner for this type of problem (i.e., scattering) is the *Boundary Element Method* or BEM. The BEM represents a surface as a mesh of discrete elements. There are two different computation methods which are employed in the BEM—direct (or collocation) and indirect (or variational). The direct method uses pressure and normal velocity on the boundary surface as the unknowns. The indirect method uses a representation of the pressure and velocity on each side of the boundary, pressure discontinuity (pressure jumps, double layer potentials), and normal velocity (velocity jumps, single layer potential), as the unknowns. The direct technique uses a nonsymmetric matrix of equations while the indirect technique uses a symmetric set.[6] Determining which method to use is a function of problem size, computing power, and computing hardware. These issues and their dependence on solution method are described in Appendix A where the justification is made for using the indirect approach. A general description of the BEM is provided in Appendix B. The elements and their size limit the interpretation of the results and frequency range of validity.

The act of discretization converts a smooth continuous surface to a set of small planar elements (in this scheme). The size of these elements directly relates to the upper frequency bound for which the discretized surface mesh is valid. Generally it is suggested that six elements are required to accurately represent a period of an acoustic wave, although in the limit four are possible. The frequency bound is then determined by the largest element such that six element edges would equal one wavelength; $f_{\max}=c/(6\times edge_{max})$, where c is the speed of sound and $edge_{max}$ is the length of the longest element edge.

IV. BEM MODEL GENERATION

The primary component of the calculation is the computational mesh. This mesh represents the physical surface in the calculation. The goal is to obtain a mesh which resembles the physical structure to be modeled as closely as possible, while still being of a manageable size. The size of the mesh determines the speed of the calculation and other hardware requirements such as memory and available disk space (see Appendix A). The structure of the mesh, specifically the maximum element size, also determines the upper frequency for which the subsequent calculation is valid. The result of the BEM HRTF calculation is a balancing of these factors: mesh accuracy, size, homogenous discretization.

A. Head geometry acquisition

Obtaining the geometrical data of an individual was the first step in designing the boundary element model to be used in the calculation. The method used here utilized a laser scanner to obtain 3D data from the surface of the object. For the data collected, a Cyberware Head Scanner (Cyberware Color 3D Scanner #4020 RGB/PS-D) was used. The system operates using a vertical linear transmitter/receiver array which is rotated about the subject, resulting in a cylindrical data set containing the spatial information of the subject (in addition to a color map which was discarded for this research). There are a number of limitations with this system. The primary limitation of the system was that it utilized an optical laser. This requires the object to be somewhat reflective. The major problem area for this was hair. It was necessary for subjects to cover their hair with a rubber cap, providing for a reflective surface. Due to the linear design of the system, the data were restricted to line-of-sight data in a radial direction about the head. Because of this, any occluded areas were not able to be measured. The result of this was that areas such as the space behind the ears and detailed folds of the ears were seen as solid-filled. The ear canal was also seen as blocked, as no light reflected from it. This second effect readily provided a model suitable for the simulation of blocked ear canal measurements. It has been shown that once the sound approaches the ear canal (within a few mm) the coding for spatial information is complete.[7] Propagation through the ear canal provides for additional filtering (a function of the length and diameter of the car canal) but this does not provide any directional information. This measurement method is commonly used as it simplifies the measurement technique (no need for placing microphones deep into the subjects ear canal) and when not employed requires the ear canal resonance to be removed from the recording and compensated for in playback unless sound sources for listening are placed at the same position in the ear canal.

As the system relies on reflections back toward the array, surfaces that were almost perpendicular to the array are difficult to view. The primary surface where this occurred was the top of the head, which is left open, as is the bottom

of the scan as the subject continues to exist beyond the range of the system. Due to the size of the array, only the head and neck of the subject were able to be scanned. This is also a limitation as it is believed by many that the shoulders and torso (and even legs when seated) all affect the HRTF to some noticeable extent. Due to the enormous data set such a model would include, and the time necessary for such a computation (see Appendix A), these effects are not considered at this stage of the research. It was acknowledged that this is a possible deficiency but was a required assumption at this stage. Finally, the typical use for this system is for graphical applications. For this reason, most of the data processing software was designed to optimize graphical rendering, i.e., fewer elements of greater size without need for strict element connectivity. These various issues required the need for complex modifications to the mesh to be performed before any acoustic calculations could be done, as described in Sec. B below.

The result of the optical data scan was a cylindrical shell, open at both ends, which was a surface model of the subject's head and neck. The data set contained approximately 300 000 triangular elements with edges approximately 1 mm in length on average. The resolution in the vertical direction was much greater than the horizontal direction, making the elements typically elongated. Allowing for six nodes per wavelength, this model would be valid up to approximately 57 kHz, well beyond the audible frequency range. Solving this problem would also take an enormous amount of computer power, storage, and time (see Appendix A).

B. Mesh modification

Several modifications were necessary to make the mesh suitable for the HRTF calculation. To make the mesh usable by the boundary element software, it was necessary to have a closed mesh. This required closing the top and bottom of the cylindrical scan mesh and sealing any holes that may have been present due to the lack of element connectivity, i.e., two adjacent elements not being connected as duplicate coincident nodes exist, one for each element. Resolution reduction of the mesh was necessary to make the frequency range limited to the audible spectrum and the problem realistic in size. The mesh also needed to be refined in some areas where large elements existed due to the apparent elongation of many elements due to the scanning method. This also lead to the goal of creating elements more equilateral in shape, which is advantageous to the boundary element solver. As no software existed to perform the changes necessary to the mesh, several programs were written to deal with these specific tasks.[3] The final mesh, after all modifications, is shown in Fig. 1. The effect of the acquisition hardware's inability to collect data behind occluded surfaces is highlighted by the area behind the pinnae, as shown in Fig. 2.

1. Closure

Creating a closed mesh (a solid surface with no holes) from the raw data required several procedures. The first requirement was to remove any coincident nodes which from visual inspection appear to be the same node, but structurally

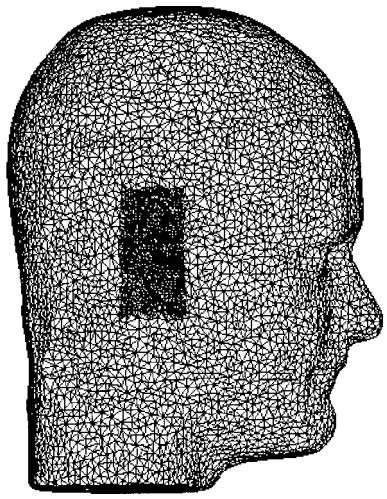


FIG. 1. Full BEM mesh used. Frequency limitation of 5.4 kHz for four nodes/wavelength. Note the increased resolution in the pinna region.

are a hole in the mesh. There was no clear method to determine which nodes were coincident with other nodes other than a brute force approach. Therefore, the distance from every node to every other node was evaluated. If the distance was too small (determined to be 0.015 cm) then the two nodes would be merged. This procedure also eliminated any elements which were very small, i.e., where an edge falls within the determined tolerance.

After resolving coincident nodes, final closure involved closing the top of the head and the neck. This was accomplished by choosing cutoff planes for both the top and the bottom at which the closure would be created. From this plane, all points either above or below were brought to the cutoff plane and then connected together. This created a plane with very long narrow elements all directed toward the center (arbitrarily chosen point). To maintain the frequency range of the model it was necessary to refine these large elements into smaller, more equilateral elements (see Sec. B 3 below).

2. Coarsen mesh

Coarsening of the mesh is the process of reducing the number of elements (and nodes) in the mesh to make the problem a manageable size, and to remove excessive resolution. Due to the nature of finite element solutions, the frequency limitation of the results is a function of the size of the

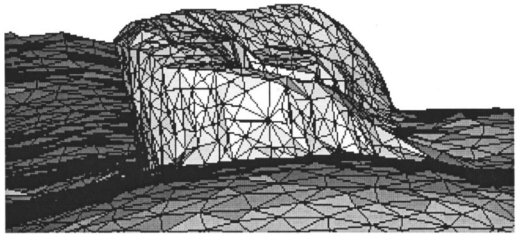


FIG. 2. Cylindrical scanner omission effect. Surfaces for which the cylindrical scanner cannot obtain a reflection from, or around, are closed or filled according to line-of-sight toward the center of the head. The effect is most prominent for the area behind the ear, which is interpreted as being filled solid.

individual elements. Particularly, the results are not valid when the largest element edge is a quarter wavelength at maximum. It is typically suggested that the edge be at least 1/6 wavelength. The raw mesh is valid (not including the closure areas at the top and bottom) up to approximately 57 kHz and has in excess of 300 000 elements. It was desired for computational reasons to reduce the mesh to no larger than 30 000 elements (see Appendix A regarding computation speed and frequency limitations). The size of an element was evaluated by determining the length of each side. If any side was smaller than the selected tolerance, that element was removed. The removal of an element is not a trivial matter in terms of connectivity and adjacent elements. It was desired that upon removing an element, there would be the minimal amount of change to the mesh in general. Two methods were considered in removing an element. The first method was to collapse the element, moving all three vertices to the center. This method removes the element in question, and the three elements which shared sides with the element in question. It was decided that this method distorted the mesh too much. The second method, the method chosen, collapses the edge in question, moving the two end points to a single midpoint. The result of this is the removal of the element in question and the single element which shared the small edge.

It was observed that if the coarsening procedure progressed through the elements sequentially, i.e., some neighboring elements were consecutively removed in the early stages of the procedure, the mesh would be corrupted and greatly distorted. This was due to the fact that early in the procedure so many elements needed to be removed. The result was a mesh that very much was a function of the numbering scheme, which should not be a factor in the mesh shape. The solution was to loop through the elements several times, skipping several elements while progressing through the mesh, each time through the loop skipping fewer elements until the final pass in which no elements were skipped. This solution solved the problem and worked quite well. This method was used for many of the procedures which affected large numbers of elements.

Since the pinna are very complicated and probably the most crucial structures in the HRTF mechanism, they were excluded from the majority of the coarsening procedures. The resulting mesh therefore contains a much greater amount of detail and number of elements in and around the pinna region, as can be seen in Fig. 1.

3. Refine mesh

The closure of the top and bottom of the head created a number of large elements. Large elements reduce the maximum frequency for which the model is valid in the solver. It was therefore necessary to refine these elements into more numerous smaller elements. A search procedure similar to the one used for coarsening the mesh was used to determine which elements needed to be refined. The length of each side of every element was calculated, and if it exceeded the maximum allowed length the element was refined. Setting the maximum length close to the minimum length used for the coarsening procedure results in the mesh being composed of somewhat regular, more equilateral, triangles. This is advantageous for the solver, in that the elements are more regularly spaced.

The method by which an element was broken into smaller elements was a matter of consideration as it was with element removal. Two possible methods for refining an element were evaluated. The first method determined the center of the element triangle. From this point, the triangle was divided into three new triangles, using the center point as a new connecting node. The difficulty with the method was that a long skinny triangle when refined in this way results in more long and skinny triangles. This is the situation that was undesired. The second method for element refining was much the same as the selected coarsening procedure. The method by which an element was refined was to first determine the midpoint of the oversized edge in question. The neighboring element which contained the same edge was then determined. Using the midpoint as a new vertex, these two elements were divided into four. This method ensured that the elements were made progressively shorter and more regular, instead of thinner while remaining long.

C. Field mesh

The field mesh is the other half of the set of points at which the transfer function was computed. For comparison purposes, the measurement mesh was generated to duplicate the measurement points for the experimentally measured HRTF, as described by Katz[8] and to be discussed in a subsequent paper. A spherical mesh was constructed which corresponded to the sphere used at Wright-Patterson Air Force Base, comprising 272 points on a 4.27 m diameter sphere.[9] Points (speakers in the real apparatus) are located at 15° increments. The BEM solution calculates the total field at a given point due to all sources. If the setup was reproduced verbatim, there would be a separate calculation for each source location on the sphere. Instead, if the properties of reciprocity are utilized, the source/receiver positions can be interchanged. The result of this switch is only two sources (i.e., two ear, two calculations) and 272 receivers. There is still some calculation due to each receiver, but these calculations are much faster than the source calculation. The initial calculation comprises solving for the potentials for all the elements such that the source and boundary conditions are met. Then, using the results of these potentials, the pressure field is calculated at each field point through superposition of all the contributions.

V. CALCULATION

A. Basic mesh

The primary mesh for the HRTF calculation is shown in Fig. 1. This paper describes results from a rigid model in which all the elements are defined as having infinite (i.e., perfectly rigid) impedance. A subsequent papers will address the effect of changing impedance characteristics of the model.

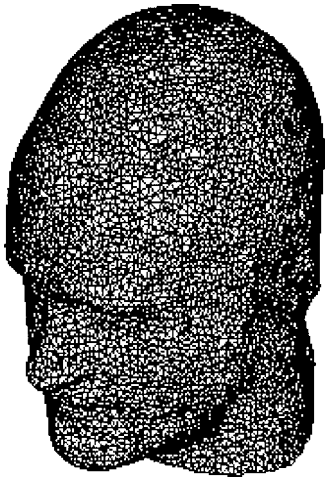


FIG. 3. BEM mesh used for evaluation of the removal of the pinna contribution.

B. Pinna contribution

As a test to see the effects of major geometric changes, a large modification was made to the measurement mesh. This modification was the entire removal of the pinna, something which would be extremely difficult with a real subject and should also prove interesting in examining the contributions of the head alone, separate from the pinnae. The pinnae were sliced off using a plane which was roughly normal to the head surface over the pinna region. The resulting mesh can be seen in Fig. 3.

C. Rigid sphere

Two spherical meshes were created and used in the HRTF calculation comparisons. The first mesh had the same diameter as the inter-aural spacing (IAS) between the entrance of the two ear canals. This sized sphere is used in some recording techniques as a simple binaural head. It provides the theoretically correct sized sphere for representing the time delay between the ears, determined by the inter-aural spacing. The mesh can be seen, in its relative position to the head mesh, in Fig. 4(a) and is referred to as either the ''small sphere'' or ''IAS Sphere.'' The same ear canal point

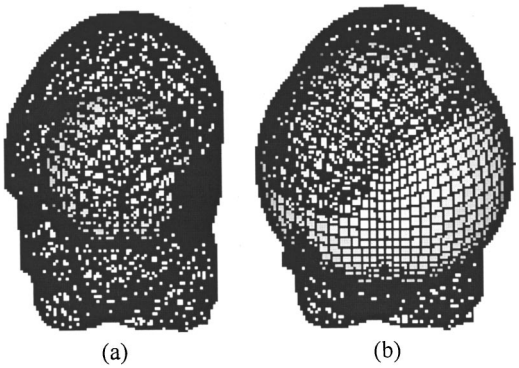


FIG. 4. BEM spherical meshes and head overlay for dimensional and placement reference. (a) IAS Sphere—size designed such that diameter of the sphere is equal to the inter-aural spacing, i.e., the edges of the sphere are coincident with the entrances to the ear canals. (b) BEM Eqvol Sphere—size designed such that the sphere and head mesh have equal volumes.

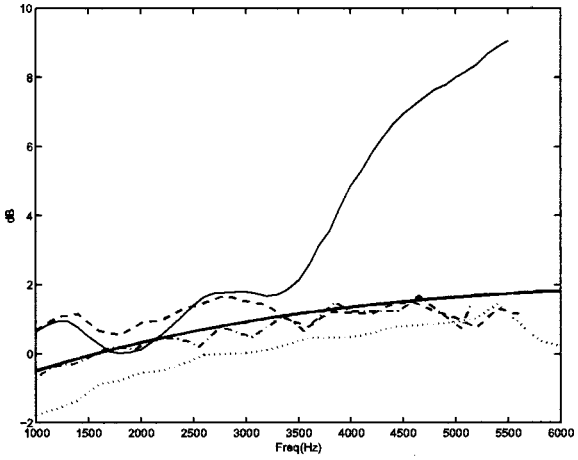


FIG. 5. Free-field equalized rigid BEM transfer functions for left ear, source directly left. (——) Head mesh, (— —) Pinnaless head, (· · · ·) IAS Sphere, (–·–) Eqvol Sphere, (—— thick) Eqvol Analytical.

(relative to the placement of the measurement mesh) was used and the transfer function was calculated.

The second spherical mesh, shown in Fig. 4(b), had the same volume as the head mesh and is referred to as the ''Eqvol Sphere.'' The selection of this mesh was to provide a sphere which could provide a better approximation of the shadowing effect of the head due to its size, at the expense of the slightly exaggerated inter-aural spacing. The distance along the surface of the sphere from ear-to-ear is more similar to that of the head mesh than the IAS Sphere, providing a more reasonable estimation of the inter-aural time delay. The mesh is orientated such that the ear canal positions are on the same axis as the head mesh. Due to the extremely high degree of symmetry of these models the results can be slightly unstable, as discussed in Appendix B. While efforts were made to minimize these effects, some errors still persist and can be seen in the results as sharp peaks or nulls. These are more present in the Eqvol Sphere as it is larger, thereby reducing the frequencies of the internal resonances more towards the region of interest. The scattering from a sphere has a well defined solution.[10,11] Results of the analytical solution for the Eqvol Sphere are presented with the first set of calculation results (see Fig. 5).

VI. CALCULATION RESULTS

Variations in the geometry of the mesh highlight contributions of the shape of the head to the HRTF. The four mesh geometries used here—full head, head without pinna, IAS sphere, Eqvol sphere—show several major points. As the HRTF is a function of space as well as frequency, different source locations may be affected differently by geometrical changes. For presentation and analysis of these results, the data were free-field equalized.

To reduce computation time in this study calculations are performed for the left ear alone. Several locations from the 272 calculated are chosen to provide a global understanding of the results. Source locations presented begin with the source and receiver on the same side of the head on the inter-aural axis. The source location is then moved from the receiver ear side around to the opposite ear side, providing

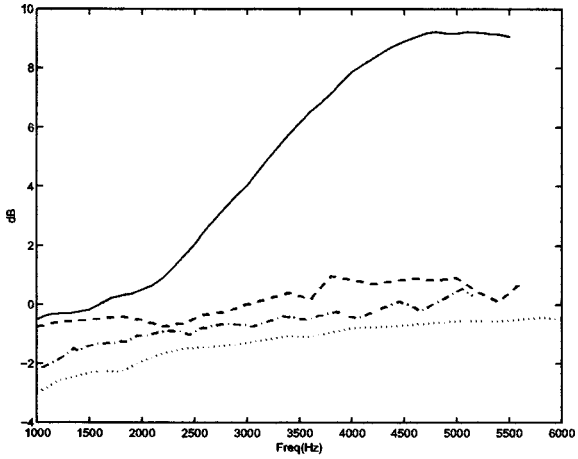


FIG. 6. Free-field equalized rigid BEM transfer functions for left ear, source front-left. (——) Head mesh, (— —) Pinnaless head, (· · · ·) IAS Sphere, (–·–) Eqvol Sphere.

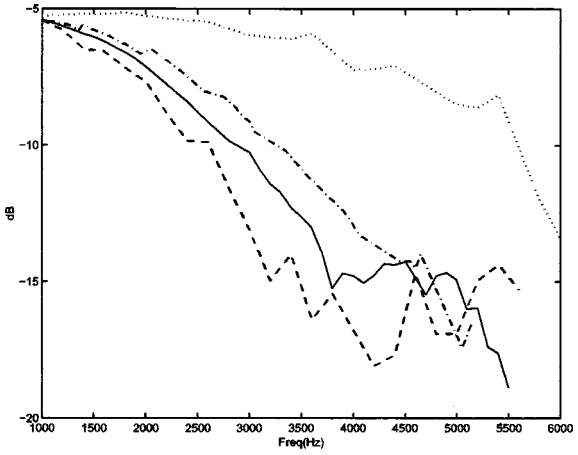


FIG. 8. Free-field equalized rigid BEM transfer functions for left ear, source directly right. (——) Head mesh, (— —) Pinnaless head, (· · · ·) IAS Sphere, (–·–) Eqvol Sphere.

an impression of the variations in the HRTF due to the various model conditions. Following this, several locations in the median plane are presented to represent elevation effects. Finally, a position off the horizontal and vertical plane is presented, showing a marked difference between model conditions.

The first position for consideration is directly left. Results for all four geometries, as well as the analytical solution to the Eqvol Sphere, are given in Fig. 5. From this result, it can be seen that the pinnaless head follows the same response as the intact head up to approximately 3 kHz, at which point it follows the trend of the eqvol sphere. The IAS Sphere response has roughly the same shape as the Eqvol Sphere but with a lower magnitude. At higher frequencies the two spheres and the pinnaless head are approaching similar values. The source position is moved clockwise in the horizontal plane to the midpoint between left and front. For this position (results shown in Fig. 6), deviation between the pinna and pinnaless model occurs at 2 kHz, and even though the pinnaless and sphere models follow the same trend, the pinnaless model is offset with approximately a 2 dB increase from the IAS Sphere and 1 dB from the Eqvol Sphere.

With the source position directly in front, the variations between the pinna and pinnaless meshes increases, as shown in Fig. 7. Trend agreement only exists up to approximately 1.5 kHz, with a 1 dB offset. The pinnaless solution follows the rough trend of the spherical solutions up to about 4.5 kHz, at which point the solutions diverge. As described in the previous section, there are frequencies at which the internal modal resonances of the mesh induce error and affect the exterior solution. For this reason, it is important to ignore sharp spikes in the BEM solution, such as the spikes in the Eqvol spherical solution in Fig. 7. Continuation of source rotation to the opposite side of the listening ear, directly right, is shown in Fig. 8. It is quite evident that the pinna, pinnaless, and Eqvol mesh solutions agree well over most of the frequency range. There are some deviations on the order of 3 dB, but these are at locations where the transfer function result is already on the order of −15 dB. The pinna and pinnaless model both have a ''step'' as a function of frequency starting in the neighborhood of 3–3.5 kHz. This is not present in the Eqvol Sphere. The total reduction is much less than for the IAS Spherical model. The final position in the horizontal plane under consideration is the rear position,

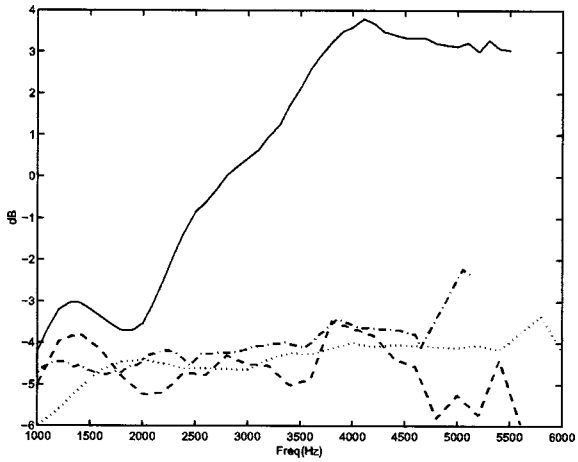


FIG. 7. Free-field equalized rigid BEM transfer functions for left ear, source directly front. (——) Head mesh, (— — —) Pinnaless head, (· · · ·) IAS Sphere, (–·–) Eqvol Sphere.

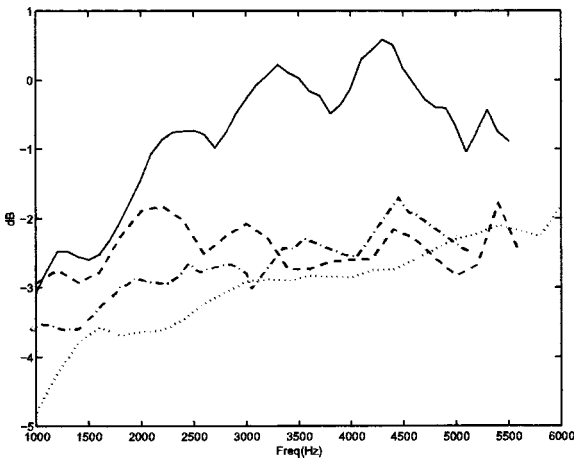


FIG. 9. Free-field equalized rigid BEM transfer functions for left ear, source directly rear. (——) Head mesh, (— —) Pinnaless head, (· · · ·) IAS Sphere, (–·–) Eqvol Sphere.

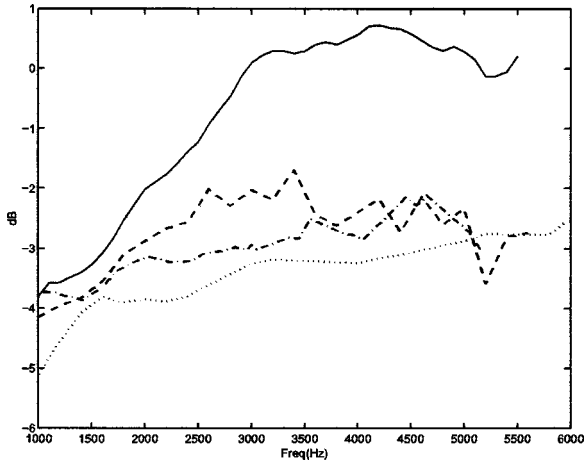


FIG. 10. Free-field equalized rigid BEM transfer functions for left ear, source median plane, elevation −45°. (——) Head mesh, (— —) Pinnaless head, (· · · ·) IAS Sphere, (–·–) Eqvol Sphere.

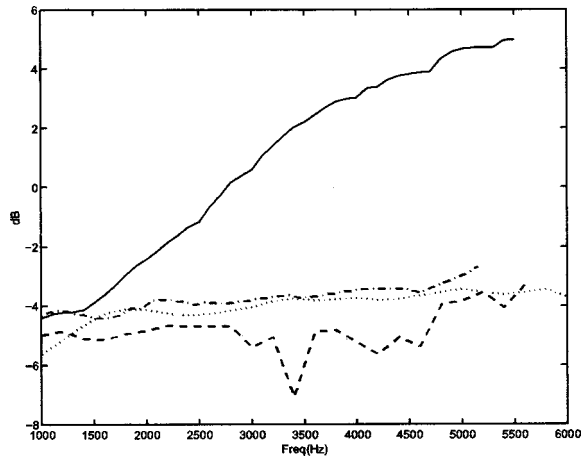


FIG. 11. Free-field equalized rigid BEM transfer functions for left ear, source median plane, elevation +45°. (——) Head mesh, (— —) Pinnaless head, (· · · ·) IAS Sphere, (–·–) Eqvol Sphere.

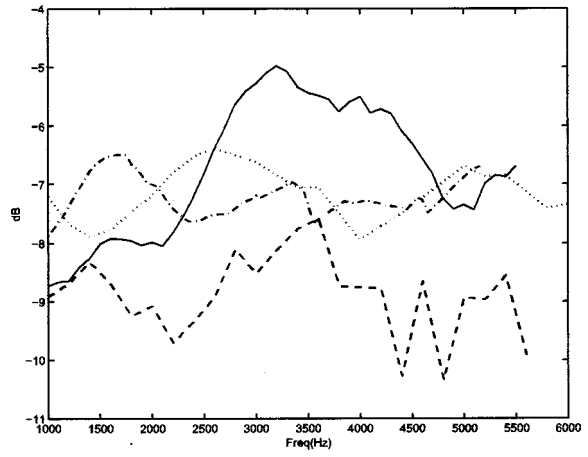


FIG. 12. Free-field equalized rigid BEM transfer functions for left ear, source center of front lower right octant. (——) Head mesh, (— —) Pinnaless head, (· · · ·) IAS Sphere, (–·–) Eqvol Sphere.

shown in Fig. 9. As with the front source (Fig. 7), the pinna and pinnaless models agree only up to about 2 kHz, at which point the pinnaless model tends more toward the spherical models. The Eqvol Sphere matches better with the pinnaless head than the IAS Sphere. Also of interest is the radical difference between the front and rear source positions for the intact model, and how the pinnaless and spherical models are not dramatically different. Two median plane sources are shown in Fig. 10 and Fig. 11. In general the pinnaless model still conforms more to the spherical models than the intact pinna model. Finally, a point not on either the horizontal or median plane is shown. In Fig. 12 the results for a source in the front sphere, in the center of the lower right octant, is shown. What is useful in reviewing these positions is the radical difference between all four models.

VII. CONCLUSIONS

It can be concluded from these results that the shape of the head in general is important to the low frequency region of the HRTF. The contribution of the pinna to the characteristics of the HRTF was distinctly shown by utilizing the computer model. With this technique the pinna of an individual subject were totally removed, leaving a plane at surface on the sides of the head. The calculated HRTFs from this computational experiment clearly identified the contribution of the pinna to the HRTF as a function of frequency and direction. For sources toward the front and rear of the listener the pinna effects begin at frequencies in the region of 1.5–2 kHz. For sound sources directed toward the ear pinna contributions begin at frequencies above 3.5 kHz. This coincides with the region reported in the literature where localization ability exists not as a function phase difference between the ears but as a function of frequency magnitude variations due to incident angle. Approximating the head with a sphere of equal volume yields very similar results to a head *without* pinna, but does not match the effects of a head with pinna except in the case of the source on the opposite side of the listening ear.

A number of assumptions were made in the construction of the mesh which differ from the true individual's head and body geometry. Primarily, these assumptions are the lack of any body below the neck. In addition was the effect of the cylindrical scanner. The following paper in this series of work will address the effect of impedance conditions on measurements and the comparison of calculated and measured HRTFs.

ACKNOWLEDGMENTS

The author is grateful to his supervisor Dr. Jiri Tichy for his assistance in this work. Further thanks are offered to Dr. Victor Sparrow, Dr. Courtney Burroughs, and Dr. Timothy Leishman for their assistance and providing insight and encouragement throughout this work. Special thanks are offered to Rodney Korte for his patience and valued assistance. Part of this work was financed by the Applied Research Lab and the Acoustics Department of The Pennsylvania State University. Sysnoise, a product of LMS Numerical Technologies (LMS-NIT), versions 5.1–5.3 were used during this re-

search. Sysnoise for the SGI was graciously donated for this research project and the author would like to acknowledge the kind contributions of LMS-NIT in aiding this extremely computationally intensive work. The author would also like to thank Kathleen Robinette, of Wright-Patterson Air Force Base, for the use and assistance with the laser scanner.

APPENDIX A: CALCULATION SPEED

Boundary element computations are very powerful, but they are also quite computationally intensive. The nature of the BEM is the solution of a set of simultaneous equations. Each equation contains a component for every element in the mesh, and there is an equation for each element. Therefore, as the size of the mesh grows, the size of the calculation grows as the square of the number of elements. There are some variations to the solution methods which help speed up the calculations, and each commercial package has their own tricks, but the general rule still applies. Sysnoise was used as the BEM solver for all the HRTF calculations here.

When discussing the ''size'' of the calculation there are really two separate issues involved. First is the amount of calculation that needs to be performed, a larger problem would need more time. The second consideration is the actual size of the set of simultaneous equations, as it refers to computer memory and disk storage space. As can be expected, as the mesh gets larger, the amount of computational space increases as well as the amount of computational time. The two variations on the BEM solution optimize each of these, but not both. (For details regarding the BEM options, see Sec. III and Appendix B.)

For problems using more than a just few hundred elements the direct method is optimized for speed. It solves a set of full simultaneous equations and in doing so is rather fast. But, in creating such a large set of equations, more storage space for the equations is required. In contrast, the indirect approach utilizes a symmetrical set of equations. The indirect method is slower than the direct method, but requires less memory (either RAM or temporary disk storage) than the direct method.[6]

In preparing for the HRTF calculation procedure using the BEM solution, several tests were made to determine the expected amount of storage space and computational time that would be required for such a problem. The required temporary storage space for the calculation (*not* for storage of the results) was examined as a function of mesh size for both the direct and indirect method. The results of this comparisons are shown in Fig. 13. This storage space can be accommodated either in RAM or on disk. As few machines have the large amounts of memory required for these calculation, the results are shown for disk space requirements.

Several different machines were available which provided for a variety of calculation speed comparisons. The results of some of these tests are given in Fig. 14. The direct method could not be tested using all the sample meshes as once the mesh size increased beyond approximately 12 000 elements, the disk space requirement for the direct calculation exceeded the free space available (1.4 Gigabytes).

It can be seen quite clearly that the computational speed quickly becomes impractical with increasing mesh size. Ideally, the computation would calculate the HRTF over the entire audible frequency range. Calculating up to 20 kHz at a high frequency resolution, for example in 25 Hz increments, would equate to 800 frequencies. The largest mesh attempted, again due to storage limitations, was 22 000 elements. This mesh, shown in Fig. 1, was only valid up to 5.4 kHz. This calculation took approximately 28 hours of CPU time for a single frequency.

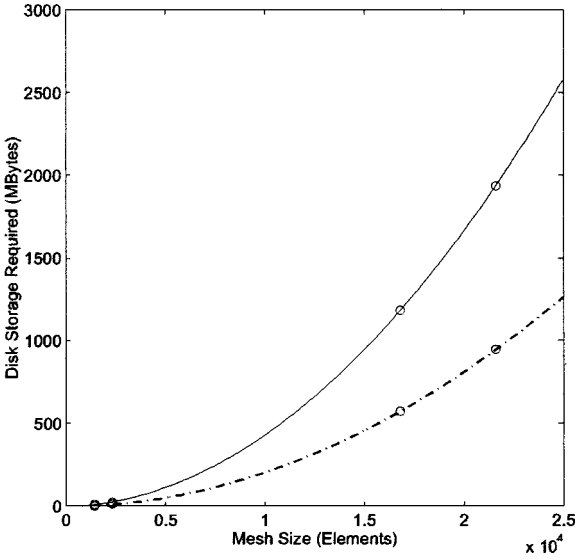


FIG. 13. Required temporary storage space for BEM calculations as a function of mesh size for (——) direct and (–·–) indirect method.

To calculate the ideal number of frequencies at this speed, neglecting the issue that at higher frequency resolution the computation times increases, would take approximately 2.6 years of CPU time. In addition, it is necessary to separately calculate the HRTF for each ear, doubling the calculation time to over 5 years. It was therefore necessary to balance the desired results with a plausible time frame. A calculation speed of one frequency/day was decided as acceptable, limiting the frequency range to 5.4 kHz. With this

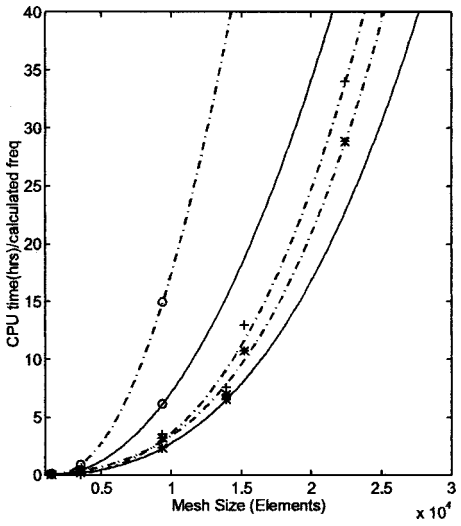


FIG. 14. Calculation speed comparison as a function of mesh size for (——) direct and (–·–) indirect calculation methods on several machines, (○) HP9000/600, (∗) DEC Alpha 3000/600, and (+) SGI Onyx R4000 (×3).

limited frequency range it would not be plausible to perform any subjective testing as the audio quality and information contained over this range would not be sufficient. As there is minimal to no frequency relevant information in the HRTF below 1 kHz,[1] the span of the calculation can be limited to the range of 1–5.4 kHz. With the removal of subjective testing, the calculation can then be limited to a single ear. This quickly reduced the computation time in half. Finally, sufficient frequency resolution was determined to be at 100 Hz intervals maximum. For some of the mesh variation conditions this spacing was increased up to 400 Hz. The result was a computation of approximately 50 days of CPU time for a single model condition.

APPENDIX B: BEM OVERVIEW

The indirect approach incorporates the interior and exterior of the boundary mesh. The indirect method can be described in the following way. (The description of the BEM method presented in Appendix B is primarily derived from the Sysnoise Manual.[6]) Using the exterior side normal, n, to define direction one can define p^+ and u^+ as the pressure and velocity on the exterior side of the surface and p^- and u^- as the pressure and velocity on the interior side. The normal derivatives of the respective pressures are $\delta p^+/\delta n$ and $\delta p^-/\delta n$. The indirect method utilizes pressure and velocity jumps. It is therefore necessary to make the following definitions where μ is the double layer potential and σ is the single layer potential:

$$\mu = p^+ - p^-, \tag{B1}$$

$$\sigma = \frac{\delta p^+}{\delta n} - \frac{\delta p^-}{\delta n} = -i\rho\omega(u^+ - u^-). \tag{B2}$$

In order to solve the set of equations which result in the values for the layer potential sources, the boundary conditions on the surface must be defined. The boundary condition can be defined by describing the pressure on the surface, normal velocity on the surface, or normal impedance on the surface.

The discretization of the surface and the layer potentials can be regarded as a monopole (single layer) and dipole (double layer) distribution which, when given the appropriate values reproduces the scattered field by the boundary ''surface.'' The acoustical variables in the fluid (i.e., pressure, velocity, intensity) can be determined by summing, through superposition, the solution set of representative sources. The pressure at a point away from the surface can be calculated using Eq. (B3), where $\oint_s$ is the surface integral over the boundary surface s, $\overline{x}$ is the field point, $\overline{x_s}$ is the source location, and $G(\overline{x}|\overline{x_s})$ is the free space Green's function:

$$\hat{p}(\overline{x}) = \oint_s \left[\mu(\overline{x_s}) \frac{\delta G(\overline{x}|\overline{x_s})}{\delta n} - \sigma(\overline{x_s}) G(\overline{x}|\overline{x_s}) \right] ds. \tag{B3}$$

In the BEM there exists the potential for errors regarding the exterior problem where the solution can be nonunique at some frequencies. The error arises at resonance frequencies of the internal problem. These frequencies, in the BEM, are referred to as irregular frequencies in the solution. To reduce the contribution of the internal resonances the methodology is to prescribe finite impedances on a set of elements (depending in solution option) within the interior of the surface. For the indirect method these are referred to as overdetermination elements (overdetermination points for the direct method). The elements are intended to suppress the internal resonances of the model. The number and placement of these elements is a function of the geometry of the mesh. An unsymmetrical mesh requires few overdetermination elements to suppress the internal resonances, while a very symmetric mesh may require many internal elements in order to suppress all the of these irregular frequencies. More details regarding this, and comparisons between solutions of a sphere and BEM predictions can be found in Katz.[3]

[1] Jens Blauert, *Spatial Hearing*, 3rd ed. (The MIT Press, Massachusetts, 1997).

[2] Brian F. G. Katz, ''Computational model of an individual head-related transfer function using the BEM,'' Acta Acoustica **85**, S267 (1999).

[3] Brian F. G. Katz, ''Measurement and Calculation of Individual Head-Related Transfer Functions Using a Boundary Element Model Including the Measurement and Effect of Skin and Hair Impedance,'' Doctoral thesis, The Pennsylvania State University, 1998.

[4] Søren Gert Weinrich, ''Sound field calculations around the human head,'' Report No. 37, Acoustics Laboratory, Technical University of Denmark, 1984.

[5] Fukudome Kimitoshi and Masaki Yamada, ''Influence of the shape and size of a dummyhead upon Thévenin acoustic impedance and Thévenin pressure,'' J. Acoust. Soc. Jpn. (E) **10**, 11–22 (1989).

[6] *Sysnoise User's Manual: Revision 5.3*, EMS Numerical Technologies, Leuven, Belgium, 1997.

[7] Dorte Hammershoi and Henrik Moller, ''Sound transmission to and within the human ear canal,'' J. Acoust. Soc. Am. **100**, 408–427 (1996).

[8] Brian F. G. Katz, ''New approach for obtaining individualized head-related transfer functions,'' J. Acoust. Soc. Am. **100**, 2609 (1996).

[9] Durand R. Begault, *3-D Sound for Virtual Reality and Multimedia* (Academic, New York, 1994).

[10] Philip M. Morse and K. Uno Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1986).

[11] Eugen Skudrzyk, *The Foundations of Acoustics* (Springer-Verlag, New York, 1971).

Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements

Brian F. G. Katz[a)]
Laboratoire d'Acoustique Musicale, Université de Paris 6, 11 rue de Lourmel, 75015 Paris, France

(Received 31 January 2001; revised 10 July 2001; accepted 20 August 2001)

Following previous work by the author involving the calculation of an individual head-related transfer function (HRTF) using a Boundary Element Method (BEM) approach, impedance conditions are now included to take account of the acoustic properties of human hair. In addition, comparisons are made here between calculations and measured values for the HRTF of a specific individual. Numerous works have been published regarding the measurement of HRTFs, but rarely are several methods compared for the same individual. The results presented in the work compare two different measurement techniques and a computational BEM for acquiring an individual HRTF. Impedance effects have been incorporated in the BEM model for hair based on measured data, providing a final set for comparison. Measurement results show significant variations for the same individual. Computational results show good agreement within the range of experimental variations. Definite trends are observed for many directions, while the limitations of the methods are also highlighted for others. The effects of incorporating hair impedance are shown to provide an improvement in the correlation of computational results, indicating an affect of hair impedance on the HRTF. [DOI: 10.1121/1.1412441]

PACS numbers: 43.64.Bt, 43.66.Qp, 43.66.Pn, 43.64.Ha [LHC]

I. INTRODUCTION

The previous work by the author has shown the initial results of limited attempts to calculate an individual head-related transfer function (HRTF) based solely on geometrical data of the head and pinnae using a Boundary Element Method (BEM) approach.[1] Earlier results utilized a rigid boundary condition model, neglecting any possible contributions to the HRTF from the impedance effects of human skin and hair. This paper incorporates recent measurements of the acoustic properties of human skin and hair to provide a complex boundary condition model for the HRTF calculation. Following this, comparisons are made with actual HRTF measurements of a given individual and the computational results with and without impedance effects.

The intent of this work is to show the feasibility and suitability of the BEM as a tool for predicting and therefore understanding the HRTF. With the ability to modify the geometry of an individual using a computer model, and to investigate the effects of these geometrical changes on the HRTF, a powerful route to comprehension of the HRTF is opened.[2] While limitations exist with the current work, these limitations highlight the avenues for further work in the area and provide a solid basis for subsequent assumptions in such work.

[a)]This work was performed while at the Acoustics Department, The Pennsylvania State University, State College, PA 16801. Electronic mail: bfgkatz@attglobal.net

II. IMPEDANCE CONDITION

A. Incorporation into BEM model

Recent work by Katz has provided measurements of the acoustic properties of human skin and hair in the audible frequency range.[3] The results of this work showed that human skin is, for all practical purposes, acoustically rigid. On the contrary, human hair was found to have a significant amount of acoustic absorption. For this reason, the acoustic impedance properties of human hair have been incorporated in the boundary conditions of the BEM model.

Using the impedance measurements of hair from Katz, the boundary conditions of some elements were modified from purely rigid to the measured impedance values as a function of frequency. As the BEM is designed around surfaces, the effects of sound propagation *through* hair is not feasible without defining a secondary fluid over part of the head having some defined impedance. Using the method here, each element in the defined region is assigned the impedance condition and it is thought that as the sound diffracts around the head, it will interact with the impedance effects of the elements it propagates over. The BEM solver used only provides for the definition of normal acoustic impedance, not grazing incidence values, and considers each element as a locally reacting surface not capable of imparting vibrations to adjoining elements. This, in combination with the boundary definition itself, makes this method for inclusion of hair impedance a rough approximation. For these reasons, the hair region is only simply defined, not precisely defined by the geometry of the original subject. The set of elements defined with the hair impedance can be seen in Fig. 1.

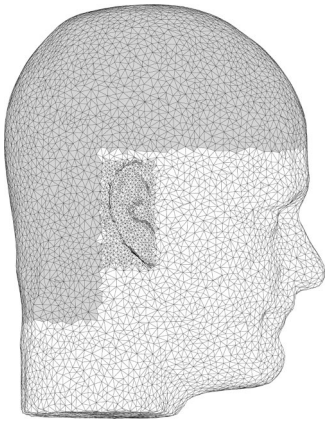


FIG. 1. Full BEM mesh used showing hair region where elements have been defined as having the corresponding impedance from measurements. Note the increased resolution in the pinna region.

B. Results

Results show the effect of hair on the intact head. Of the 272 source locations calculated, results are shown for the same locations as used by Katz.[1] Using the left ear as the receiver and placing the source directly to the left provides the data given in Fig. 2(a). As can be seen, in the region above 3 kHz there is approximately a 0.5–1 dB decrease in the transfer function due to the presence of hair. Below 3 kHz there is a general increase of roughly the same order, or greater (1.5 dB). Variations of such a small degree, although interesting, are difficult to attribute either a great deal of meaning to or consider as generally perceivable to the listener. As the source is moved clockwise in the horizontal plane, toward the front, the effect of the hair changes slightly as shown in Fig. 2(b). At frequencies above 4.5 kHz there is an increase due to the hair of 1 dB. In the range below this there is little contribution due to the hair. Much the same hair variation is found when the source is at the front of the head model [Fig. 2(c)]. Moving the source to the opposite side of the listening ear shows some real effect between the models, as shown in Fig. 2(d). Variations in the region above 3.5 kHz begin at about 1 dB and increase to over 6 dB, where the response due to hair is greater than without. Below this frequency, the hair has the effect of reducing the transfer function by up to 3 dB. The rear source position shows similar results for the head model, though not to the same degree [Fig. 3(a)]. The response shows an increase for the hair model beginning at approximately 3 kHz, though only reaching a maximum deviation of 1 dB. This highlights the fact that the HRTF and its contributing geometrical parameters are greatly a function of angle. A change which assists the model in one location may have little, or the opposite, effect at other incident angles. The lower left position as well as the median plane source positions [shown in Fig. 3(b) thru Fig. 3(d)] show similar results as seen for the other front sources. Variations on the order of 1 dB or less exist. For the low median source below approximately 3 kHz the hair response is lower, but above 3 kHz the hair response is greater by the same degree. This is interesting, although with the lack of a torso in the model, it is difficult to attribute much meaning to this occurrence at this time. The torso provides additional

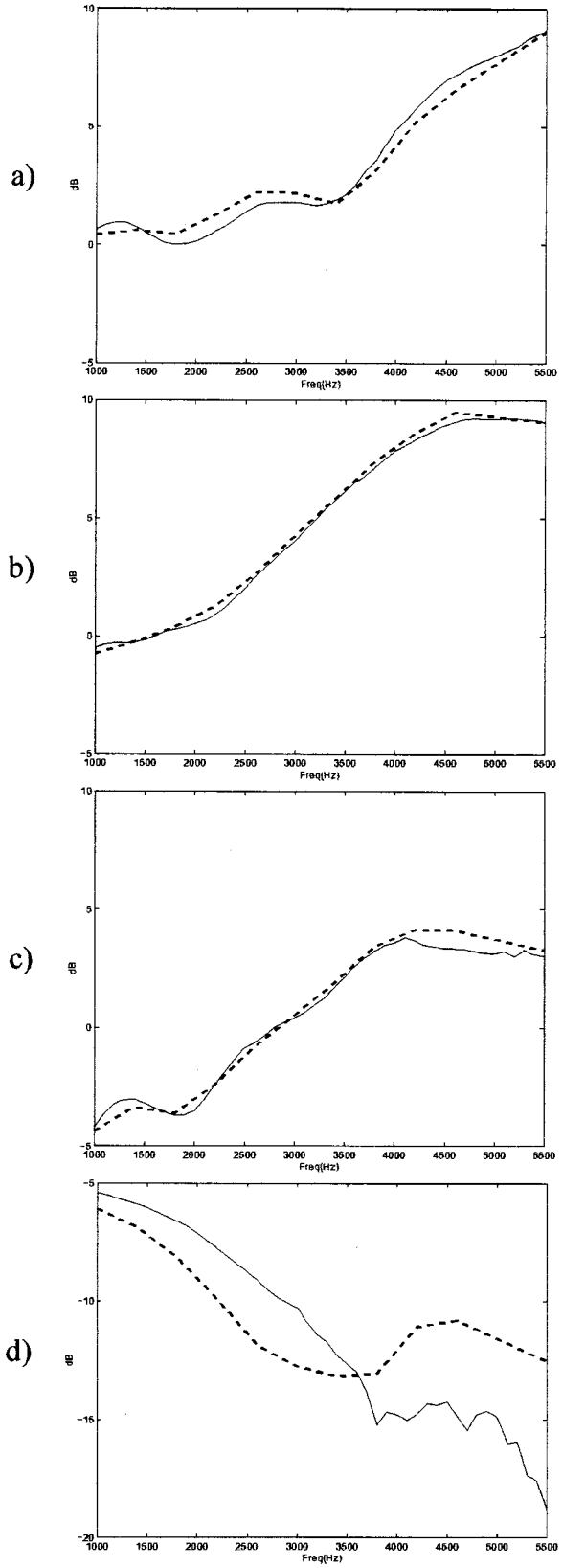


FIG. 2. Free-field equalized BEM transfer functions for left ear. (a) Source directly left, (b) source front-left, (c) source front, (d) source right. (——) Rigid head mesh, (––) Hair impedance boundary condition.

reflections from the chest, back, and shoulders. In addition, it provides a kind of baffle which blocks sound from going under the head, which can occur in the BEM model. From these calculations it can be seen that the inclusion of hair

effects (limited by the definition mechanism used) does result is some degree of variation in the HRTF. These variations range from 0.5 dB to 6 dB for the source positions shown. Whether or not this inclusion of hair improves the model as compared to measured data is discussed in the Sec. IV.

III. HRTF MEASUREMENTS

Two techniques for measurement of individual HRTFs were employed for this study. Both tests utilized the blocked ear canal measurement method, which has been shown to be sufficient for directional information.[4] This method is most commonly used for HRTF measurements where listening simulations are utilized. Measurements made near the eardrum contain ear canal resonances of the subject which must be removed before being listened to unless the reproduction system is also located at the eardrum. Measurement data are taken from Katz.[2,5] Listening tests have not been incorporated into this study.

The equipment used for this measurement was located at the WPAFB Armstrong Laboratory in Dayton, Ohio. The measurement apparatus consisted of a spherical array of sources, a pair of receivers, and appropriate hardware. The source array was a dome frame, 4.27 m in diameter, containing 272 speakers with 10 cm (4 in.) diameter. The speakers were located at 15° increments. The dome was located inside an anechoic chamber on a suspended floor (see photo in Ref. 6, p. 143).

The tubular frame construction of the dome was covered with foam to reduce reflections. A pair of receiving electret microphones, one for each ear, were each mounted on a small plastic stick. Foam ear plugs were modified to contain a tube in the center. When the foam plugs were inserted into the ear canal, the microphone stick was placed in the tube, thereby mounting the microphone in the center of the ear canal, flush with the end of the foam plug providing for a blocked ear canal measurement. The data acquisition was performed using a spectrum analyzer. The noise source used was a random white noise signal, and data was taken with 100 averages. The resulting data consisted of a left and right ear power spectrum for each source position, and an interaural time delay (ITD) data set which contained the frequency averaged time delay between the two ears for each source position.

Subsequent to acquisition of the HRTF a calibration measurement was made. This measurement used a rotating apparatus which was placed in the position previously occupied by the head (or subject). The same two sensor microphones were placed on a rotating boom. Each microphone was directed toward each individual source on the sphere, and the transfer function was measured. This transfer function, when used as a calibration function, removed the effects of the electronics, source, and receiver. This resulted in a transfer function which contained the effect of just the presence of the head.

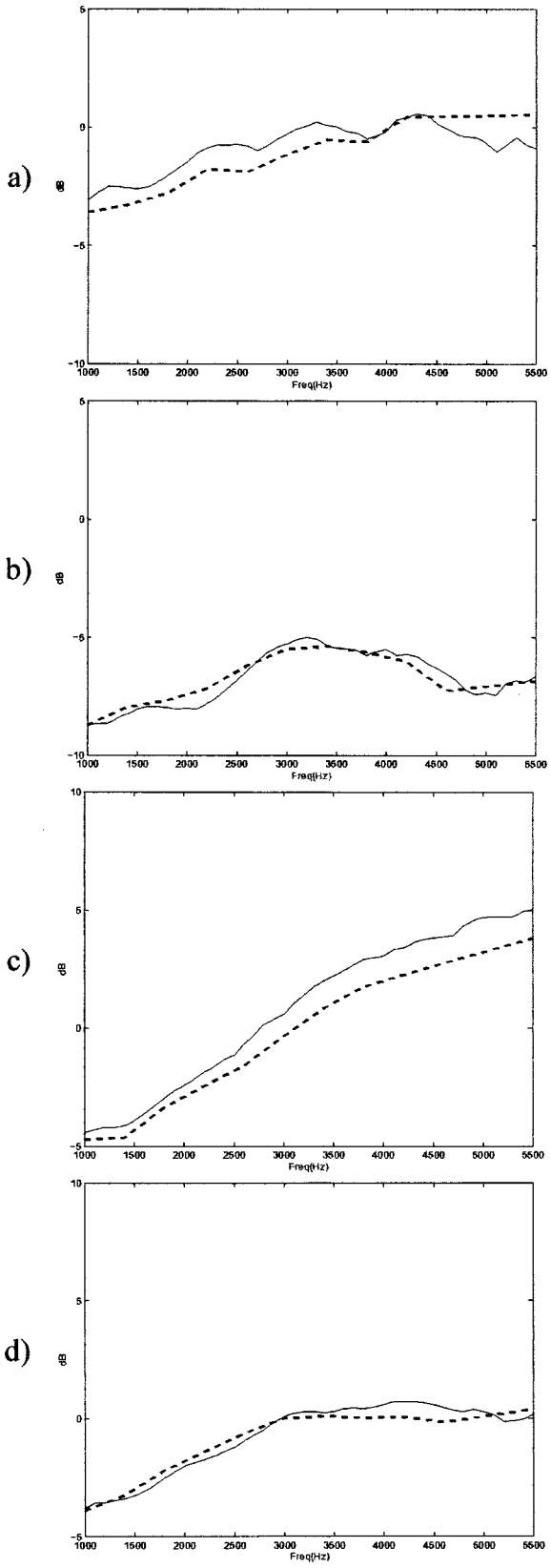


FIG. 3. Free-field equalized BEM transfer functions for left ear. (a) Source rear, (b) source front lower right octant, (c) source front median plane +45°, (d) source front median plane −45°. (——) Rigid head mesh, (– –) Hair impedance boundary condition.

The measurement of an individual HRTF in this facility was a very time consuming experiment. For this reason, the laboratory was experimenting with using pinna replicas, made from real individuals, from which the HRTF measurement was made. This method was hoped to provide a sufficient HRTF measurement for use in localization research. This measurement method provided for an interesting comparison between individual (with real impedance conditions) and semi-individual (with rigid impedance conditions) HRTF measurements for this study.

A. Real individual HRTF measurements

For the individual HRTF measurement, the subjects were seated inside the dome. The head was positioned so that the center of the interaural axis was located at the center of the sphere. This was accomplished using two low powered lasers to align the head correctly.

The subject's location was stabilized using a chin rest, thereby keeping the head reasonably immobile during the measurement. This chin rest frame was covered with acoustically absorptive foam in order to reduce any effects to the HRTF. Allowing for settling time and data averaging, this procedure lasted about 45 min. This time included several rest periods for the subjects. Due to the time involved in this measurement, only a subset of the entire source array was used. Out of the 272, only 70 source positions were measured. These positions consisted of the horizontal plane, the median plane, and three points in the middle of each of the eight octants. This provided ample data for comparison. Data measured with this method are labeled *Real*.

B. Pinnae replica semi-individual HRTF measurements

Due to the long measurement time for an individual HRTF measurement, an alternate method was employed using pinna replicas made from the individual subjects and placed onto a mannequin head. Use of this method requires several assumptions. As the mannequin head is the same for all subjects, it is assumed that the individual specific information required for localization is contained solely in the effects due to the pinna. Any differences in head shape are ignored.

It is also assumed that the surface impedance of the pinna replica and mannequin head are either comparable to an actual subject or are of no consequence to the HRTF. The validity of these assumptions is dealt with both through comparisons between the measured data sets and in varying the impedances of the computational model (as seen in Sec. II).

The mannequin used for the pinna replica measurements was the Bio-acoustic Anthropomorphic Manikin for Auditory Research (BAMAR) head built at WPAFB. Material composition of the head consisted of an epoxy fiberglass mixture which resulted in a very rigid and solid structure. The head was comprised of three sections, allowing for the interaural spacing (distance between the ears) to be adjusted to that of the subject whose pinna are being used. This allowed for the interaural time delay (ITD) to be correct for the HRTF measurement. The gaps between the sections were filled with clay to create an acoustically closed surface. Finally, the head is placed onto a torso which simulates the effect of the human torso in terms of reflections from the chest, back, and shoulders. The chin rest from the individual HRTF measurements was removed, as was the chair used by the subject.

Pinna replicas were built by taking a casting of the subject's pinna and surrounding area. The ear canals were first blocked with a cotton plug. The final pinna molds were composed of 40% Dow Corning 3110 RTV compound using Type I catalyst, and 60% Dow Corning Silicone Fluid Type 200 [viscosity 10 000 cs]. Having created the pinna molds, they were then placed into position on the dummy head. The microphones were placed in the pinna molds in the same manner and position as for the individual HRTF measurements. With the mannequin in place in the measurement chamber, the transfer function was measured for all 272 positions. As there is no human subject in the test chamber, the signal levels can be greatly increased from the real subject HRTF measurement, providing a much greater signal to noise ratio. Data measured with this method are labeled *Mannequin*.

IV. CALCULATED AND MEASURED INDIVIDUAL HRTF RESULTS

Final analysis of the work presented here consists of comparing the HRTFs obtained through both experimental methods and calculations, both with and without hair. All HRTF measurements and BEM calculations we done using the same individual. Data are shown for the left ear only.

One major difference between the computational models and the experimentally measured models is the presence, or lack thereof, of the shoulders and torso. Previous works have discussed the effects of reflections from the shoulders and torso on the HRTF.[6–9] Torso contributions are on the order of ±3–4 dB in the region from 100–250 Hz up to 500 Hz–2 kHz, depending on the study. Shoulder effects have been reported as variations of ±2–5 dB in the ranges of 800 Hz–1.2 kHz and 2–3 kHz. A subsequent study by Searle *et al.* showed minimal contributions to localization performance due to shoulder reflections, far less than head and pinnae effects.[10]

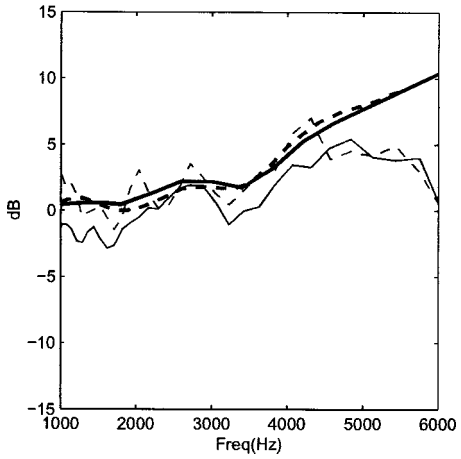


FIG. 4. Full HRTF set comparison for source left. BEM results in bold: (––) Rigid head mesh, (——) Hair impedance boundary condition. Measurements: (––) Mannequin, (——) Real.

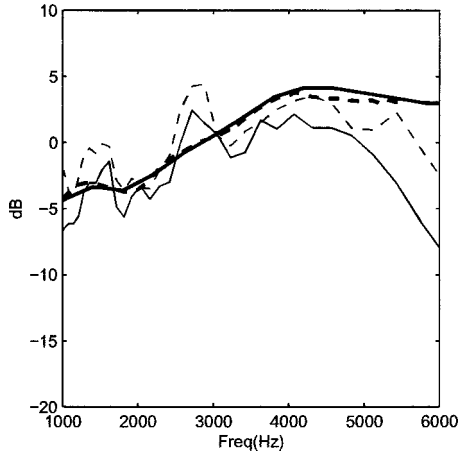


FIG. 5. Full HRTF set comparison for source front. BEM results in bold: (––) Rigid head mesh, (——) Hair impedance boundary condition. Measurements: (––) Mannequin, (——) Real.

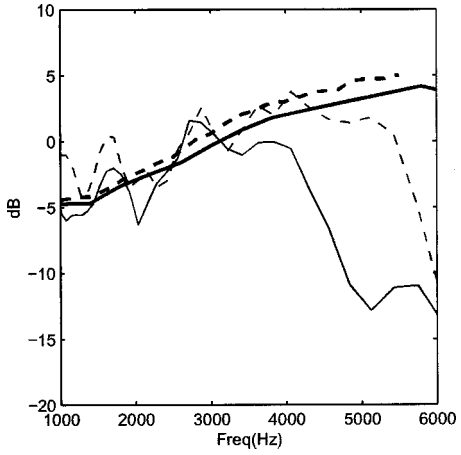


FIG. 7. Full HRTF set comparison for high frontal median plane source, +45°. BEM results in bold: (––) Rigid head mesh, (——) Hair impedance boundary condition. Measurements: (––) Mannequin, (——) Real.

The first position for consideration is source directly left. Results for the two measurement methods and the two calculation conditions are given in Fig. 4. Variations between the measured results are on the order of 2–3 dB, with the Mannequin results being slight higher magnitude up to 4.5 kHz. This could be due to the different absorption characteristics of the subject (clothed, hair) as differed from the mannequin. All four data sets follow very similar trends up to 4.5 kHz, at which point the calculated results continue to rise while the measured data decreases. The cause for this disparity is unclear. At lower frequencies, below 2.5 kHz, there are fluctuations in both measured data results which are not present in the calculation. This could likely be due to reflection effects which are attributed to this region.

Figure 5 shows the results for the source located in front. Again, all results follow similar trends with the measured results containing more fluctuations at lower frequencies. Two major peaks are present in both measured data sets, at 1.5 and 3 kHz. These are most likely due to torso effects which are most prominent for sources located in front. There is a steeper drop off at high frequencies for the real measured versus the mannequin data. This could again be attributed to the higher absorptive characteristics of the real subject.

The source right position results are shown in Fig. 6. Differences between impedance conditions of the BEM results, which are greatest for this source location, are reflected in the real measurements. Values are greater for the rigid model case in the region of 1.5–3 kHz and are lower at higher frequencies. The measured results contain many more fluctuations, though mostly on the order of 1–2 dB, than the calculated results. Even though, the calculations follow the same trend, with the inclusion of the drop then rise near 3.75 kHz.

Figure 7 shows results for a source in the median plane at +45°. Measured results agree only up to 3.25 kHz, with the real data decreasing more rapidly than the mannequin data, though they appear to be approaching the same values above 6 kHz. Correlation between measured and computed results can be seen below 3–4 kHz, but again there are some prominent peaks which are not present in the calculations. As mentioned above, these are attributed to the shoulder/torso effects which would be rather prominent for an elevated source. A low elevation source is presented in Fig. 8. Agreement between all results is rather good, with deviations being

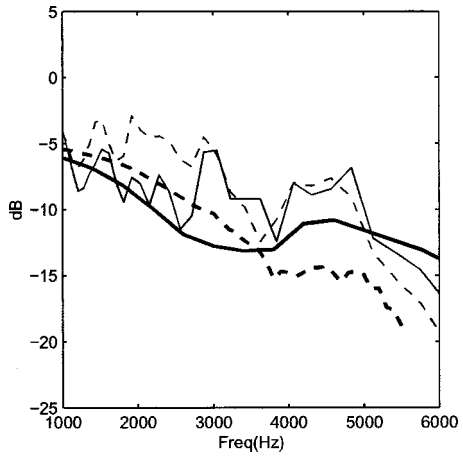


FIG. 6. Full HRTF set comparison for source right. BEM results in bold: (––) Rigid head mesh, (——) Hair impedance boundary condition. Measurements: (––) Mannequin, (——) Real.

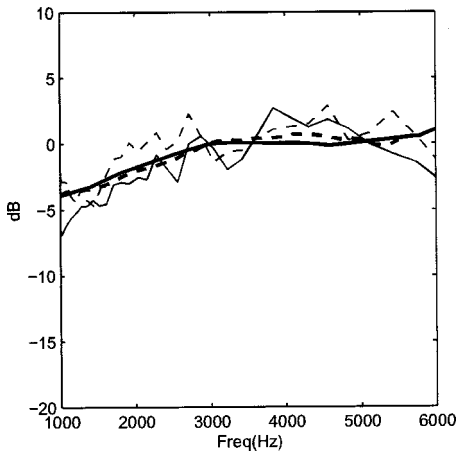


FIG. 8. Full HRTF set comparison for low frontal median plane source, −45°. BEM results in bold: (––) Rigid head mesh, (——) Hair impedance boundary condition. Measurements: (––) Mannequin, (——) Real.

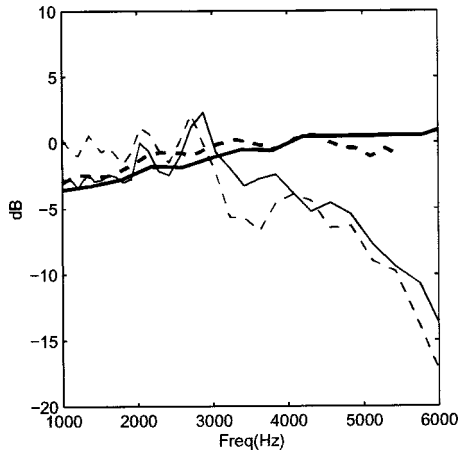


FIG. 9. Full HRTF set comparison for rear source. BEM results in bold: (––) Rigid head mesh, (——) Hair impedance boundary condition. Measurements: (––) Mannequin, (——) Real.

no greater than a few dB over the range of 1–6 kHz.

The final configuration shown is for the source position in the rear, shown in Fig. 9. There is general trend agreement up to 2.5–3 kHz, after which the measured HRTF decreases while the calculation remains constant. Similar peaks exist in the measured data as seen for other source locations. The potential for differences between the measured and calculated HRTF is rather great for this source location as the geometry acquisition system was not capable of modeling the rear surface of the pinnae, but instead resulted in a filled region (see discussion in the previous paper[1]). This would be more noticeable at higher frequencies, which is what has been observed.

V. CONCLUSIONS

This work has focused primarily on the first part of this complex procedure, the Head-Related Transfer Function. A computational model based on the geometric shape of an individual was used to calculate a limited frequency HRTF. This model showed good agreement with the various measured data, within the observed variations of the experimental results. Previous work has been shown that measurement of the HRTF is not a trivial matter (see, for example, Refs. 4, 7, 11–16 and 6, 17 for an extensive list of references). Variations between measurement methods persists and must be taken into consideration when comparing measurements and making statements regarding the physics of the system. The use of replacement individuals, utilizing personal pinna replicas in a generically shaped head, has shown relatively good agreement with measurements made on the respective real individual. Future work involving the detailed analysis of the contribution of measurement technique, experimental error, and repeatability is recommended to fully understand the variations observed.

The success of this work now provides a useful tool in examining the contribution of head and pinna shape to the properties of the HRTF. With the resource intensive nature of the BEM, further work could be limited to just modeling the pinnae in a baffle. This would allow for more detail to be considered with greater elements and a higher frequency limit while still maintaining a problem of manageable size, making it possible that a full frequency band HRTF spanning the entire audible frequency range could be computed. Other methods for geometry acquisition could also be used which account for occluded surfaces.

Hair has been found to have significant absorptive properties, increasing as a function of frequency. The contribution of acoustical parameters of the individual, the impedance of hair, on the HRTF were examined physically using an experimental rigid head with replica pinna and numerically by introducing the measured impedance data into the HRTF computation. Both procedures showed nontrivial variations in the HRTF. The effects were more pronounced, though not limited to, sources originating at or behind the ears in direction. The most prominent effect was for the sound source directed at the ear on the opposite side of the head. This is an effect in contrast to the pinnae effect which diminished for sources on the opposite side of the head.[3] Results here showed differences on the order of 6 dB due to the inclusion of hair in the head model. Future work in this area could involve the inclusion of hair in the computational model through a different approach, allowing for propagation through the material, and handling of grazing incidence conditions. By modeling the hair as an absorptive medium, through which the sound must propagate, instead of the current boundary surface representation, the results may be a more accurate representation of the actual physical phenomenon.

In the continuing path of research into human localization ability, it is hoped that this research has provided some new information, and may lead to further investigations using the same techniques. The history of this field is long, and yet, there is still much to be learned in what some perceive as a relatively new field. With advances in hardware miniaturization, computational speed, and signal processing techniques, the possibilities of research in this area continue to grow rapidly.

ACKNOWLEDGMENTS

The author is grateful to his supervisor Dr. Jiri Tichy for his assistance in this work. Further thanks are offered to Dr. Victor Sparrow, Dr. Courtney Burroughs, and Dr. Timothy Leishman for their assistance and providing insight and encouragement throughout this work. Special thanks are offered to Rodney Korte for his patience and valued assistance. Part of this work was financed by the Applied Research Lab and the Acoustics Department of The Pennsylvania State University. Sysnoise, a product of LMS Numerical Technologies (LMS-NIT), versions 5.1–5.3 were used during this research. Sysnoise for the SGI was graciously donated for this research project and the author would like to acknowledge the kind contributions of LMS-NIT in aiding this extremely computationally intensive work. The author would also like to thank Richard McKinley for the use of the BioAcoustics Lab's facilities of Wright-Patterson Air Force Base and the assistance of Dennis Allen, David Ovenshire, and Robert Bolia with HRTF acoustic data collection. Thanks are also offered to Scott Foster and Martin Walsh for their assistance with additional HRTF measurements.

[1] Brian F. G. Katz, ''BEM calculation of individual HRTF. I. Rigid model calculation,'' J. Acoust. Soc. Am. **110**, 2440–2448 (2001).
[2] Brian F. G. Katz,''Measurement and Calculation of Individual Head-Related Transfer Functions Using a Boundary Element Model Including the Measurement and Effect of Skin and Hair Impedance,'' Doctoral thesis, The Pennsylvania State University, 1998.
[3] Brian F. G. Katz, ''Acoustic absorption measurement of human hair and skin within the audible frequency range,'' J. Acoust. Soc. Am. **108**, 2238–2242 (2000).
[4] Dorte Hammershoi and Henrik Moller, ''Sound transmission to and within the human ear canal,'' J. Acoust. Soc. Am. **100**, 408–427 (1996).
[5] Brian F. G. Katz, ''New approach for obtaining individualized head-related transfer functions,'' J. Acoust. Soc. Am. **100**, 2609 (1996).
[6] Durand R. Begault, *3-D Sound for Virtual Reality and Multimedia* (Academic, New York, 1994).
[7] M. B. Gardner, ''Some monaural and binaural facets of median plane localization,'' J. Acoust. Soc. Am. **54**, 1489–1495 (1973).
[8] K. Genuit, ''A Model for the Description of Outer-Ear Transmission Characteristics'' [in German], Doctoral thesis, Rhenish-Westphalian Technical University, Aachen, Germany, 1984.
[9] H. W. Gierlich, ''The application of binaural technology,'' Appl. Acoust. **36**, 219–243 (1992).
[10] C. L. Searle, L. D. Braida, M. F. Davis, and H. S. Colburn, ''Model for auditory localization,'' J. Acoust. Soc. Am. **60**, 1164–1175 (1976).
[11] E. A. G. Shaw and R. Teranshi, ''Sound pressure generated in an external-ear replica and real human ears by a nearby point source,'' J. Acoust. Soc. Am. **44**, 240–249 (1968).
[12] E. A. G. Shaw and M. M. Vaillancourt, ''Transformation of the sound-pressure from the free field to the eardrum presented in numerical form,'' J. Acoust. Soc. Am. **78**, 1120–1123 (1985).
[13] R. A. Butler and K. Belendiuk, ''Spectral cues utilized in the localization of sound in the median sagittal plane,'' J. Acoust. Soc. Am. **61**, 1264–1269 (1977).
[14] J. Hebrank and D. Wright, ''Spectral cues used in the localization of sound sources in the median plane,'' J. Acoust. Soc. Am. **56**, 1829–1834 (1974).
[15] Fred Wightman and Doris Kistler, ''Headphone simulation of free field listening. I: Stimulus synthesis,'' J. Acoust. Soc. Am. **85**, 858–867 (1989).
[16] Fred Wightman and Doris Kistler, ''Headphone simulation of free field listening. II: Psychophysical validation,'' J. Acoust. Soc. Am. **85**, 868–878 (1989).
[17] Jens Blauert, *Spatial Hearing*, 3rd ed. (The MIT Press, Massachusetts, 1997).

Distortion-product otoacoustic emissions measured at high frequencies in humans[a)]

Laura E. Dreisbach[b)] and Jonathan H. Siegel
Department of Communication Sciences and Disorders, Northwestern University, 2299 North Campus Drive, Evanston, Illinois 60208

(Received 13 April 2001; revised 6 August 2001; accepted 7 August 2001)

Distortion-product otoacoustic emissions (DPOAEs) elicited with stimulus frequencies less than or equal to 8 kHz have been used in hearing clinics to assess whether the middle ear and cochlea are normal, but high-frequency hearing (>4 kHz) is most vulnerable to cochlear pathology. It might prove useful to measure DPOAEs with even higher frequency stimuli (>8 kHz), but there have been few reports of such studies in humans. DPOAEs have been measured in other mammals to the upper range of hearing sensitivity. The purpose of this study was to compare some characteristics of DPOAEs in human subjects elicited with high-frequency stimuli with those that have been extensively measured with lower-frequency stimuli. The primary goal was to establish if the same phenomenon responsible for the behavior of low-frequency DPOAEs is responsible for the behavior of high-frequency DPOAEs. Specifically, the DPOAE level with stimuli varied from 2 to 20 kHz, growth functions of DPOAEs, effects of varying the primary frequency ratio (f_2/f_1) on the DPOAE level, and DPOAE group delay were determined. Because the behaviors appeared to vary smoothly with stimulus frequency, the study suggests that emissions measured from 2 to 20 kHz were the product of the same biological process.
[DOI: 10.1121/1.1406497]

PACS numbers: 43.64.Jb, 43.64.Yp, 43.64.Ha [BLM]

I. INTRODUCTION

Using a microphone and speaker sealed into the ear canal of human subjects, Kemp (1978) demonstrated that under certain circumstances the cochlea not only receives sound but also produces recordable acoustic energy, called otoacoustic emissions. Otoacoustic emissions can be generated spontaneously or in response to a variety of acoustic stimuli. When the acoustic stimulus consists of two continuous pure tones of different frequencies presented simultaneously, acoustic distortion products at frequencies not present in the acoustic stimuli are produced by intermodulation between the primary stimulus tones, and can be detected in the ear canal. Due to nonlinear interaction between the primary tones, distortion occurs when the traveling waves of the lower frequency primary tone (f_1) and the higher frequency primary tone (f_2) overlap along the basilar membrane. The distortion-product frequencies are arithmetic combinations of the frequencies of the primary tones. In humans, the most prominent and widely studied distortion-product otoacoustic emission (DPOAE) occurs at the frequency $2f_1-f_2$.

Distortion-product otoacoustic emissions (DPOAEs) generated by low-level stimuli have been shown to be physiologically vulnerable to the same factors that affect or destroy hair cells, as seen in animal studies (Kim, 1980; Kim *et al.*, 1980; Lonsbury-Martin *et al.*, 1987; Brown *et al.*, 1989). Brown *et al.* (1989) reported that outer hair cell damage, confirmed by histological examination, corresponded to reduced DPOAE levels, supporting the involvement of outer hair cells in the generation of DPOAEs. Basilar-membrane tuning properties also depend on the status of hair cells (Khanna and Leonard, 1986; Ruggero and Rich, 1991), as does the sensitivity and frequency selectivity observed in the tips of neural tuning curves (Liberman and Dodds, 1984).

DPOAEs have been extensively studied and characterized in animals (e.g., Wiederhold *et al.*, 1986; Brown, 1987; Lonsbury-Martin *et al.*, 1987; Whitehead *et al.*, 1990) and for frequencies less than or equal to 8 kHz in humans (e.g., Harris *et al.*, 1989; Gaskill and Brown, 1990; O Mahoney and Kemp, 1995; Stover, Gorga *et al.*, 1996). DPOAE testing has become a quick, objective test used in hearing clinics to determine the status of the cochlea and middle ear. The majority of DPOAE studies in humans has used low-frequency stimuli (typically$\leqslant$8 kHz), but hearing at higher frequencies is most vulnerable to cochlear insults. Often, sensorineural hearing loss appears first and becomes more severe at high frequencies and then progresses to lower frequencies (see Ohlemiller and Siegel, 1994). Siegel and Hirohata (1994b) have demonstrated that high-frequency DPOAEs can be measured in humans. Thus, these measures could be related to characteristics of the mechanics of the basal portion of the human cochlea and they might lead to a new objective

[a)]Portions of this work were presented in "The effect of ear canal acoustics on high frequency distortion-product otoacoustic emission group delay measurements," Proceedings of International Evoked Response Audiometry Study Group Conference, Lyon, France, August, 1995, "Further characterization of ultra-high frequency distortion-product otoacoustic emissions (DPOAEs)," Proceedings of Association for Research in Otolaryngology, St. Petersburg, Florida, February, 1996, and "Growth functions of $2f_1-f_2$ distortion-product otoacoustic emissions are similar for stimulus frequencies above and below 10 kHz," Proceedings of Association for Research in Otolaryngology, St. Petersburg, Florida, February, 1997.

[b)]Now at San Diego State University, San Diego, CA; electronic mail: ldreisba@mail.sdsu.edu

method for the early detection of high-frequency hearing loss.

Given the uncertainty of the mechanism of the cochlear amplifier, which is thought to generate emissions, we questioned whether a single mechanism could account for the behavior of otoacoustic emissions over the entire range of hearing. Somatic motility (Dallos and Evans, 1995a, b), and hair bundle motor (Martin *et al.*, 2000) mechanisms have been proposed, but not proven, as the basis of the cochlear amplifier. Without characterizing the behavior of high-frequency DPOAEs with varied stimulus parameters, we cannot be confident that the same phenomenon responsible for the behavior of low-frequency DPOAEs is responsible for the behavior of high-frequency DPOAEs. For example, because human DPOAEs are generally lower in level than those of other mammals, nonlinearity in the middle ear might contribute significantly to the intermodulation distortion we measure in humans for high frequencies. Because no physical system is perfectly linear, the middle ear must generate some intermodulation distortion. Under most conditions where otoacoustic emissions have been studied, the distortion generated in the cochlea is dominant, at least in normal ears. There is also evidence for considerable differences in cochlear mechanics in the apex and base (Ohlemiller and Siegel, 1994; Cooper and Rhode, 1997) The experiments described in this paper characterized the behavior of DPOAEs extending the range of stimuli to higher frequencies than those used in previous studies in humans.

The levels of DPOAEs with stimuli ranging from 2 kHz to the upper limits of human hearing were first measured, using a more accurate method of calibration than that used in the Siegel and Hirohata (1994b) study (see Sec. II). Then, three additional experiments were conducted to characterize the emission behavior.

DPOAE growth functions, with varied levels of the two pure-tone stimuli, were completed to characterize the level dependence of DPOAEs. Reported DPOAE level input/output (I/O) functions, with a fixed level difference between the primary frequencies, differ between studies. Stover, Gorga *et al.* (1996) reported varied overall shapes of DPOAE I/O functions across f_2 frequency, whereas Harris and Probst (1990) demonstrated linear DPOAE functions as the levels of the primary frequencies were increased together, with increasing slopes as the stimulus frequencies increased. When the level of one primary frequency is fixed and the other is varied, the DPOAE level has been found to be more dependent on L_1 than L_2. This relationship is dependent on the frequency region examined and the levels of the stimulus tones (Whitehead, Stagner *et al.*, 1995). The maximum DPOAE level occurs with L_1 greater than L_2, for L_1 of 65 dB SPL or lower (Gaskill and Brown, 1990; Whitehead, Stagner *et al.*, 1995). For L_1 greater than 65 dB SPL, equal-level stimulus tones result in the largest DPOAE level (Whitehead, Stagner *et al.*, 1995).

Next, we examined the effects of varying the primary frequency ratio (f_2/f_1) on the DPOAE level. Previously, it has been reported that the DPOAE level as a function of varied primary frequency ratio results in a nonmonotonic function (Wilson, 1980; Harris *et al.*, 1989; Gaskill and Brown, 1990; Abdala, 1996; Stover *et al.*, 1999). The DPOAE level decreased at narrow and wide primary frequency ratios and increased to a maximum at intermediate ratios for each fixed value of f_2, but the peak of the function (i.e., f_2/f_1 at which the level peaked) decreased as f_2 increased.

Finally, the DPOAE group delay (the slope of the phase versus frequency curve) was measured. Generally, DPOAE group delay decreases with increasing stimulus frequency (Kimberley *et al.*, 1993; Whitehead *et al.*, 1994; Stover *et al.*, 1994; Stover, Neely, and Gorga, 1996; O Mahoney and Kemp, 1995; Moulin and Kemp, 1996; Whitehead *et al.*, 1996).

If high-frequency DPOAEs are the result of the same biological process responsible for low-frequency DPOAEs, then similar, continuous trends in the behavior of the emissions should be observed for these varied stimulus parameters throughout the range of stimulus frequencies examined.

II. METHODS

Specific subject characteristics and stimulus parameters used for each experiment will be detailed with the results of each experiment in Sec. III. The methods described below were identical for all experiments in this study.

A. Subjects

A total of 28 human subjects participated in the experiments, although only a subset of these subjects participated in individual experiments. Each subject demonstrated normal pure-tone sensitivity ($<$15 dB HL) between 0.25 and 8 kHz, based on routine clinical audiograms (ANSI, 1996). All subjects reported a negative history for significant ear disease, including noise exposure. Each subject's right ear was tested. The experiments were conducted with the subject slightly reclined in a comfortable chair within a double-walled sound isolation chamber. The tympanic membrane had to be visualized with the probe assembly positioned at the entrance to the ear canal so that a probe tube could be positioned at the eardrum for calibration purposes. If the subject's eardrum could not be visualized, he/she was not included in the study.

B. Equipment

DPOAEs and behavioral thresholds were measured with a custom-made system that allowed the optical placement of a probe tube at the eardrum with the emission probe in place. Calibration using eardrum sound-pressure levels measured at the eardrum was used to avoid calibration artifacts for stimulus frequencies above 2 to 3 kHz due to standing waves (Siegel, 1994; Dreisbach and Siegel, 1995; Siegel, 2002).

The emission probe itself was a custom-built version of the Etymotic Research ER-10 (ER-10S), with three stainless-steel tubes passing through the body (Fig. 1). One tube was used for sound delivery. The signals from two high-frequency drivers (Radio Shack super tweeters #40-1310B), as modified by Chan *et al.* (1993), were mixed passively before being output through the common sound delivery tube. Note that the sound source was not extended beyond the end of the probe assembly in the first experiment, DPOAE level as a function of stimulus frequency. This

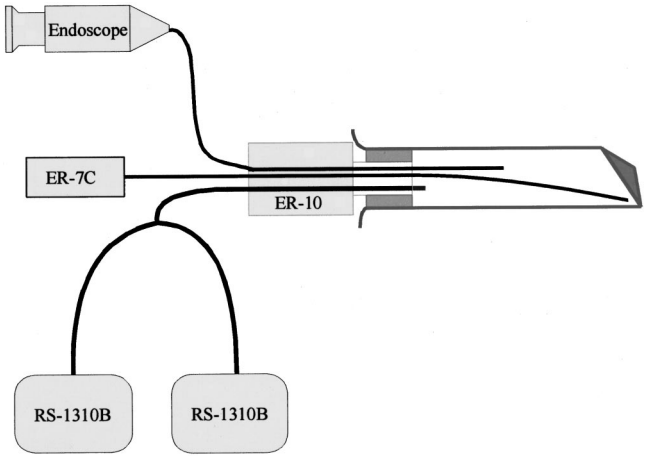


FIG. 1. Diagram of the apparatus used to collect the data in these studies. A custom-designed version of the Etymotic Research ER-10 otoacoustic emission probe (the ER-10S) allows the insertion of a 1-mm o.d. endoscope (Electro FiberOptics) into the canal entrance to observe the placement of a miniature probe tube (Etymotic Research ER-7C) near the eardrum. Stimuli were delivered by two modified Radio Shack Super Tweeters (RS-1310B).

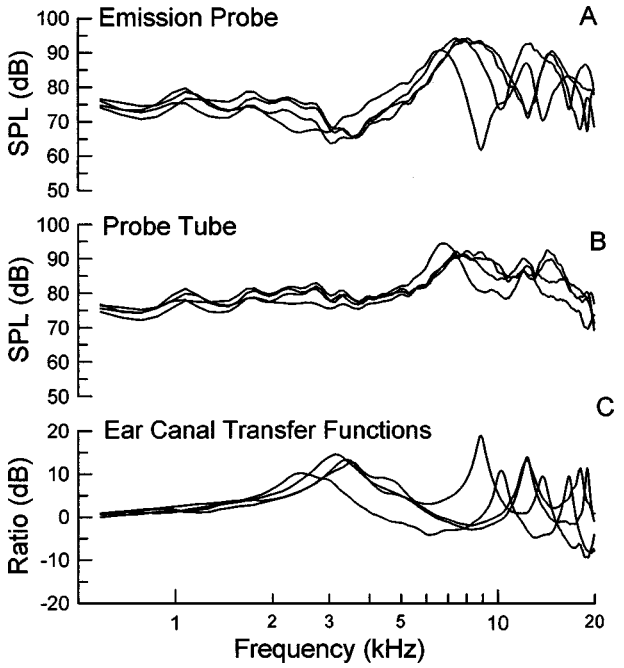


FIG. 2. The emission probe response (A), probe tube response (B), and the ear-canal transfer function (C) with optimal placement are shown for four subjects.

means that the sound source and the emission probe microphone were near one another. This can be a problem because of evanescent waves. The evanescent wave is the result of the medium immediately around a sound source moving in sympathy with the displacement of the sound source. The evanescent wave is always in phase with the source displacement. The amplitude of the evanescent wave decreases rapidly as the microphone is moved away from the sound source (Brass and Locke, 1997). Thus, effects due to evanescent waves could occur when the calibration and measurements were made using the emission probe microphone (ER-10S). However, as Brass and Locke (1997) pointed out, the errors due to evanescent waves are minimal in comparison to the errors that are made by standing waves in the occluded ear canal (Siegel, 1994). The sound source was extended beyond the end of the probe assembly for the remaining experiments, eliminating the issue of evanescent waves.

A silicone rubber probe tube (Etymotic Research ER7-14C), connected to an Etymotic Research ER-7C probe microphone, was passed through the second stainless-steel guide tube to position the probe tube near the eardrum. Evanescent waves will not affect calibration measurements made when the probe tube (ER7-14C) is placed at the eardrum, due to the distance between the sound source and the probe tube. The position of the probe tube was assessed visually with a 1-mm o.d. endoscope (Electro FiberOptics Corp.) that was inserted into the ear canal through the third stainless-steel guide tube. The endoscope formed a clear image of the eardrum and probe tube from a distance of 10–15 mm from the eardrum. The probe tube was carefully advanced under visual control until the subject reported contact with the eardrum; then, the probe tube was withdrawn slightly. During preliminary data collection, if the ear canal was touched during placement of the probe tube, subjects reported either pain or a ''pinching'' sensation. However, when the eardrum was touched the subjects reported a scratching sound or a pressure.

An ''optimal'' placement of the probe tube was considered to be one in which there were no obvious pressure nulls in the probe frequency response below 20 kHz, as seen in Fig. 2(B). Such nulls appear at the position of the probe tube due to an out-of-phase interaction between the incident pressure wave and the wave reflected from the eardrum (Siegel and Dreisbach, 1995). If quarter-wave pressure nulls were measured in the range of interest with the probe tube microphone, the probe tube was advanced so that the pressure nulls were shifted above the frequency range of interest. To illustrate this point, the emission probe and probe tube responses with optimal placement for four subjects are shown in Fig. 2. These four subjects participated in the DPOAE level as a function of stimulus frequency experiment (to be described). The emission probe data [Fig. 2(A)] show pronounced intersubject variability for frequencies above 2 kHz, due to standing waves. This variation is expected from differences in canal geometry and the depth of insertion of the emission probe. It should be noted that the variability in the probe tube responses [Fig. 2(B)] is considerably less in the same frequency range. The ear-canal transfer functions [Fig. 2(C)], the ratio between the emission probe and probe tube responses, reflect the same degree of variability seen in the emission probe measurements.

The entire assembly was stabilized by a stainless-steel rod, which extended from the back of the emission probe, supported by a modified impedance probe headband. A slight inward pressure was applied to the end of the rod. This arrangement kept relative movements of the probe and the head small, even when the subject moved. Calibration procedures were completed prior to the presentation of each set of stimuli. The stability of the measurements was excellent, as judged by the invariance of the calibrations during the entire experimental session. The calibration procedure used in these measurements differed from those of Siegel and Hi-

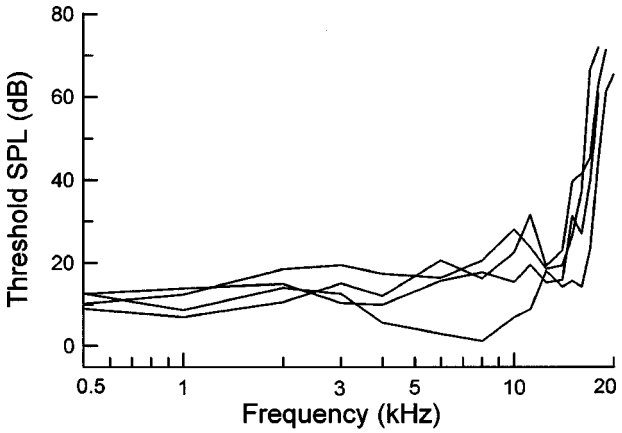


FIG. 3. Behavioral thresholds for four subjects using the same acoustic system as for emission measurements.

rohata (1994b), who used constant voltage rather than acoustic calibration.

To verify that a signal measured in the ear canal was an emission, and not an artifact of the equipment, the intermodulation distortion of the system at $2f_1-f_2$ was measured in a cavity that approximated the dimensions of the occluded ear canal. Intermodulation distortion products were found to be at least 80 dB below the stimulus levels for the stimulus conditions used in this study.

C. Procedures

Behavioral thresholds were collected using a modified Békésy tracking procedure for discrete frequencies from 0.5 to 18 kHz. The stimuli were 250-ms shaped tone bursts presented at 2/s. The computer controlled a programmable attenuator (attenuation step size was 2 dB) that was adjusted using feedback from a button controlled by the subject. The subject was instructed to hold the button down when they heard the tone and release the button when they no longer heard the tone. The investigator monitored a graph of attenuation versus time to ascertain that the subject had reached a stable tracking of threshold. Behavioral threshold was determined at each frequency by averaging the midpoint of the attenuation values in six consecutive runs of increasing intensity. Threshold measurements were referenced to eardrum sound-pressure level by measuring with the ER-7C probe tube placed near the eardrum. The reference sound-pressure level for the results of the behavioral threshold measurements is not affected by evanescent waves because the probe tube placed near the eardrum is far enough away from the sound source, as was described previously.

For all subjects, behavioral threshold curves show a steep rise above 10 kHz. To be included in the study, a subject's behavioral threshold values had to be 30 dB SPL or less for frequencies from 0.5 to 8 kHz, and the subject had to respond to frequencies up to at least 16 kHz. The thresholds at 16 kHz were within 20 to 55 dB of the subject's highest threshold for frequencies from 0.5 to 8 kHz. The variability in our subjects' high-frequency hearing thresholds is not unexpected given that hearing thresholds at high frequencies can vary greatly among subjects (Frank and Dreisbach, 1991). Behavioral threshold data are shown in Fig. 3 from four different subjects (the same subjects as in Fig. 2) who participated in the DPOAE level as a function of stimulus frequency experiment (to be described).

Emission data were collected with the EMAV (Neely and Liu, 1993) software using the Ariel DSP-16+ digital signal-processing board. The sample rate was 50 kHz. The stimulus calibration portion of the emission data collection used the probe tube microphone, positioned at the tympanic membrane. However, the input to the A/D was switched to the emission probe to collect the emission data because the probe tube microphone did not have a low enough noise floor to measure signals as small as otoacoustic emissions. Calibration using the probe tube microphone maintained constant stimulus levels at the eardrum. In addition to using the probe tube placed at the eardrum for the stimulus calibration phase of the emission data collection, the emission probe microphone was used as well. With this calibration procedure, the emission probe was used for both calibration and emission data collection. Using the emission probe microphone for calibration purposes maintained constant stimulus levels at the emission probe inlet versus the eardrum. This was done so that comparisons could be made between the traditional method of calibrating and using a probe tube placed at the eardrum.

The EMAV software samples the ear-canal sound pressure in 40-ms time windows, accumulates these samples into two interleaved buffers (sample A and sample B), and averages each buffer in the time domain. The grand average of the two samples ((A+B)/2) is fast-Fourier transformed to estimate the amplitudes and phases of both stimuli and distortion products present in the ear canal. The amplitude of the noise floor at each Fourier component, including the distortion-product frequencies, is estimated from the spectrum of the difference between the two samples (A−B). Samples are collected until one of two automatic stopping criteria have been met: the noise floor at the distortion-product frequency is less than −20 dB SPL or 4 s of artifact-free sampling have been averaged.

III. RESULTS

The figures depict data collected using the probe tube (solid lines) or emission probe (dashed lines) for calibration. Because the probe tube calibration offers better control over stimulus levels, measurements using this calibration procedure were best suited to studying trends in emission behavior over a wide range of stimulus frequencies. This is the primary objective of the study. The data using the emission probe for calibration are included because that is the conventional way of calibrating in most studies. Results obtained with the probe tube used for calibration will be presented first, followed by the results obtained with the emission probe used for calibration.

A. DPOAE level as a function of stimulus frequency

The DPOAE level across frequency was measured in four female adult subjects, ages 23–30 years. "Optimal" probe tube placements were required for these subjects because the frequency of $f_2(f_2>f_1)$ ranged from 2 to 20 kHz with a fixed stimulus frequency ratio of 1.2. Stimulus levels

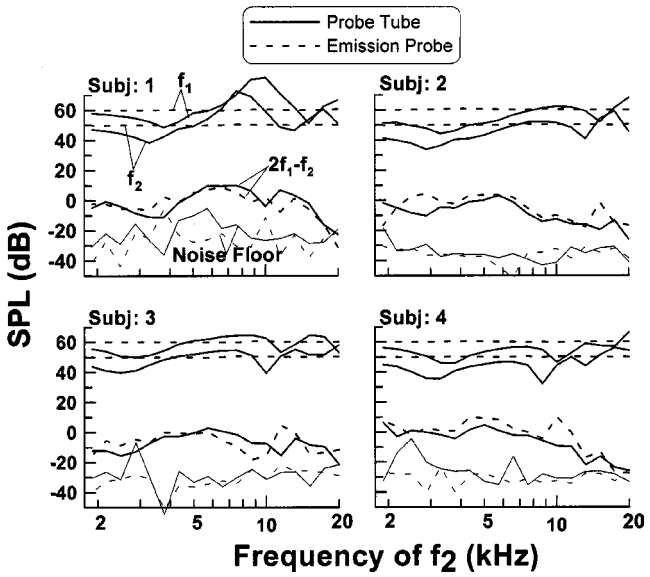


FIG. 4. Emission levels for four different subjects for f_2 frequencies from 2 to 20 kHz. The solid lines represent the data collected using the probe tube microphone for calibration and the dashed lines represent the data collected using the emission probe for calibration. The stimulus levels are indicated by the lines at the top of each panel. The stimulus levels measured by the emission probe vary with frequency (solid lines) because the eardrum sound-pressure level is maintained constant. The lines around 0 dB SPL are the 2 f_1-f_2 emission and the lines around −20 dB SPL represent the noise floor.

were 60 and 50 dB SPL for f_1 and f_2, respectively. It should be noted that, depending on which microphone was used for calibration (probe tube versus emission probe), the stimulus levels presented to the eardrum could vary from those specified.

Figure 4 shows the levels of the primary frequencies and the DPOAE across frequency. Each panel in Fig. 4 represents data collected from one subject. As can be seen in this figure, when stimulus levels were calibrated using the probe tube near the eardrum (solid lines), each subject had measurable DPOAEs up to at least 15 kHz at levels ranging from 50 to 70 dB below the level of f_1. The level of the DPOAE typically decreases for f_2 frequencies greater than 10 kHz. Note the variability of the primary levels measured by the emission probe when the probe tube microphone is used for calibration. Recall that the specified stimulus levels, 60 and 50 dB SPL, will be observed at the inlet of the microphone used for calibration, either at the probe tube near the eardrum or at the inlet of the emission probe. The data depicted in this figure were all collected using the emission probe microphone. The deviations of the measured levels of the stimuli from the target values when the probe tube is used for calibration are a direct reflection of the ear-canal transfer function between the locations of the two microphones.

Because reference input levels to the ear vary depending on which microphone (i.e., the probe tube or the emission probe) is used for calibration, measured emission levels should also depend on which microphone is used for sound calibration by the emission data collection software. Thus, to make comparisons to the data collected using the probe tube microphone for calibration, data collected using the emission probe microphone for calibration are also shown in Fig. 4 (dashed lines). The overall patterns of the DPOAE level with frequency are similar, no matter which microphone is used for calibration. However, differences in the DPOAE level of at least 10 dB are seen in all four subjects at some frequencies. The differences in the emission levels most closely tracked the deviations in the level of f_1 at the eardrum. Generally, a larger emission level was measured when the emission probe microphone was used for calibration than when the probe tube microphone was used for calibration.

To quantify the differences in emission levels due to the procedure used for calibration, the emission level at each frequency collected using the probe tube for calibration was subtracted from the emission level collected using the emission probe for calibration. For the four subjects from Fig. 4, the differences were greatest from 2.2 to 4.3 kHz and from 6.6 to 20 kHz. These are the ranges of stimulus frequencies for which quarter-wave pressure nulls were present in the emission probe microphone response for these subjects [see Fig. 2(A)]. The differences in emission levels were as great as 20 dB.

B. Growth of DPOAE as a function of stimulus level

The level dependence of DPOAEs is presumed to be related to the form of compressive mechanical nonlinearity seen in the cochlea. Studying level dependence is therefore a good way to investigate whether emissions measured in different stimulus frequency ranges share a common origin.

1. Incrementing both stimulus levels together ($L_1>L_2$ by 15 dB SPL)

Distortion-product otoacoustic emission data were obtained from six adult ears, three males and three females, with ages ranging from 22–30 years (mean =26 years). The effect of using varied unequal stimulus levels ($L_1>L_2$ by 15 dB SPL) on the 2 f_1-f_2 emission level was explored. L_1 was varied from 65 to 30 dB SPL in 3-dB steps, while L_2 was always 15 dB SPL lower than L_1. The frequency of f_2 ranged from 2 to 13 kHz (2, 2.5, 3, 4, 5, 6, 8, 10, and 13 kHz). The stimulus frequency ratio (f_2/f_1) was 1.2.

As can be seen in Fig. 5, for all f_2 frequencies, the mean distortion-product level for these subjects grows initially from the noise floor as the levels of the stimuli increase with a slope near 1 dB/dB [unity slope is indicated in Figs. 5(B) and (E)], when the probe tube microphone is used for calibration (solid lines). At the highest levels tested, the functions do not appear to be saturating for any f_2 frequency.

Notice that the emission probe (dashed line) microphone calibration yielded similar growth curves to those measured with the probe tube microphone used for calibration at 2 and 13 kHz. However, at 4, 6, 8, and 10 kHz the distortion-product level is greater for the emission probe than for the probe tube calibration procedure. Again, this is the frequency region of the first quarter-wave pressure null at the emission probe, so the stimulus levels are higher at the eardrum than at the emission probe when the emission probe is used during calibration. Additionally, saturation was only seen at the highest levels tested, for 6 kHz, when the emission probe was used for calibration. Overall, when both primaries are varied together with a fixed level difference, the main trend

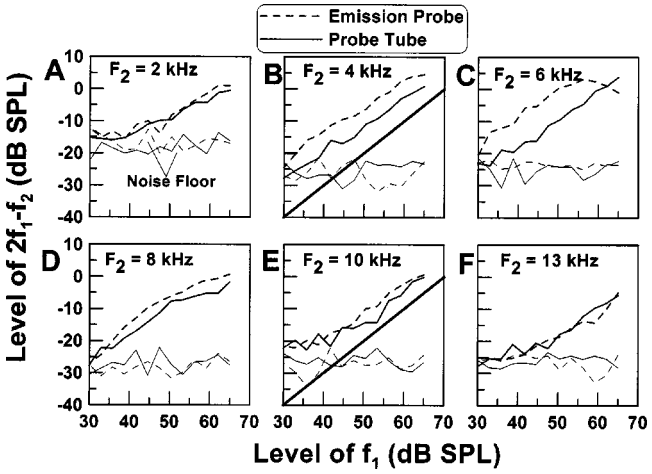


FIG. 5. Average DPOAE level measurements with $L_1 > L_2$ by 15 dB at the specified f_2 frequencies for a group of six subjects. The DPOAE level grows initially from the noise floor as the stimulus level increases with a slope near 1 dB/dB [unity slope is indicated in (B) and (E) with a dark solid line]. The dark solid lines depict the level of the DPOAE measured with the probe tube microphone used for calibration and the dark dashed lines depict the data collected when the emission probe microphone was used for calibration. The thin solid lines and dashed lines represent the noise floor measurement obtained with the probe tube microphone and the emission probe microphone, respectively, for calibration.

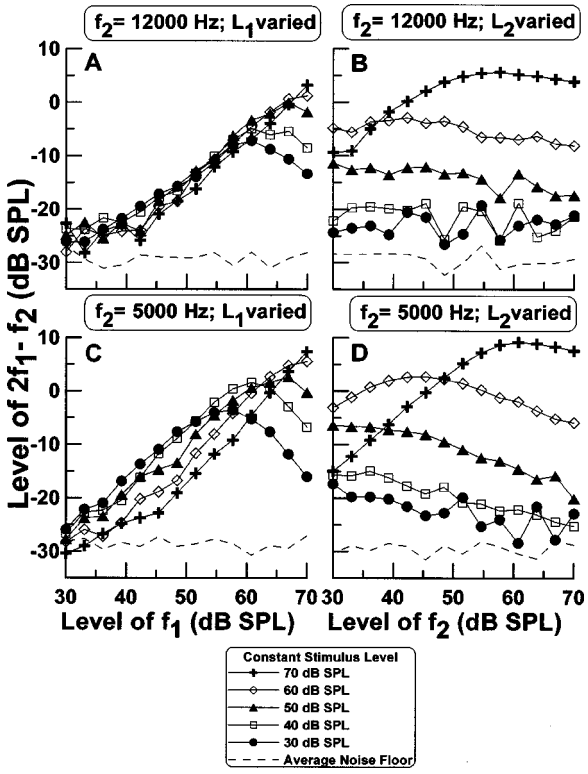


FIG. 6. Emission levels measured with varied L_1 and fixed L_2 (A) and (C) illustrate a rate of growth of around 1 dB/dB to a maximum and decline with increasing L_1 for most of the range of constant levels of L_2 for both 12 and 5 kHz, respectively. For higher fixed levels of L_2 the emission level grows monotonically over the entire range of L_1. The maximum emission level increased with higher levels of the fixed stimulus. The maximum emission level was seen when L_1 was greater than L_2. Emission levels measured with varied L_2 and fixed L_1 (B) and (D) illustrate shallower growth with varied L_2, with the exception at 70 dB SPL. There was typically little dependence on L_2 over the entire range of L_1, except when the fixed stimulus was 70 dB SPL.

is an almost linear slope for L_1 of 60 dB SPL or lower, for calibration using either the probe tube or the emission probe microphones.

2. Incrementing the level of one stimulus while the other is fixed

The effect of using varied unequal stimulus levels (L_1 varied while L_2 was fixed and vice versa) on the $2f_1-f_2$ level was explored for six f_2 frequencies (2, 5, 10, 12, 14, and 16 kHz). The stimulus frequency ratio (f_2/f_1) was 1.2. L_1 was held constant at 70, 60, 50, 40, and 30 dB SPL while L_2 was varied for each L_1 from 70 to 30 dB SPL in 3-dB steps. The procedure was repeated holding L_2 constant and varying L_1 over an identical range. Eight subjects (five female, three male) ranging in age from 22 to 32 years participated in the study. The data in Figs. 6 and 7 represent averaged responses from this population of subjects.

Published reports demonstrate generally steeper growth of $2f_1-f_2$ for increasing L_1, compared to increasing L_2 (Gaskill and Brown, 1990; Hauser and Probst, 1991; Whitehead, McCoy *et al.*, 1995). Figures 6(A) and (C) illustrate a rate of growth of around 1 dB/dB to a maximum and decline with increasing L_1 for most of the range of constant levels of L_2 for both 12 and 5 kHz, respectively. The data depicted in this figure were obtained with the probe tube microphone used as the reference for calibration. For higher fixed levels of L_2 the emission level grows monotonically over the entire range of L_1. The maximum emission level increased with higher levels of the fixed stimulus and was always reached when L_1 was greater than L_2. Figures 6(B) and (D) illustrate shallower growth with varied L_2, except when L_2 was 70 dB SPL. We presume that the emission level would have increased with increasing L_2 at lower fixed levels of L_1 had we varied L_2 below 30 dB SPL. There was typically little dependence on L_2 over the entire range of L_1, except when the fixed stimulus was 70 dB SPL. The patterns illustrated in Fig. 6 were similar for all other frequencies tested (data not shown).

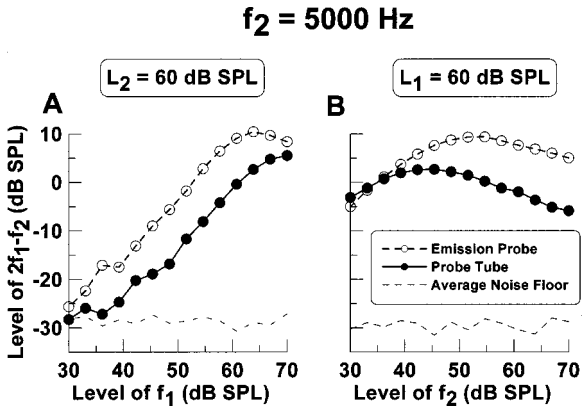


FIG. 7. DPOAE growth functions are illustrated for $f_1 = 4167$ Hz and $f_2 = 5000$ Hz with either L_2 (A) or L_1 fixed (B) (60 dB SPL), while the other primary level is varied. The solid lines with the filled symbols represent the data obtained using the probe tube microphone for calibration and the dashed lines with the open symbols depict the data collected with the emission probe microphone used for calibration. The rate of growth is steeper for increasing L_1 with L_2 fixed than for increasing L_2 with L_1 fixed. The thin dashed line in each panel represents the measured noise floor.

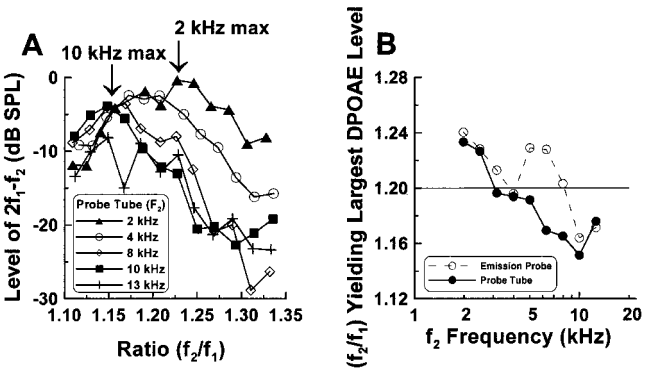


FIG. 8. (A) The mean data for eight subjects for the level of $2f_1-f_2$ as a function of stimulus frequency ratio (1.11 to 1.33) at specified f_2 frequencies with the probe tube microphone used for calibration. The stimulus frequency ratio (f_2/f_1) yielding the largest DPOAE level (peak) decreases as f_2 increases. The arrows indicate the stimulus frequency ratio yielding the maximum DPOAE level for 10 and 2 kHz. (B) The f_2/f_1 producing the maximum DPOAE level for fixed f_2 frequencies from 2 to 13 kHz. The solid and dashed lines represent the data collected when the probe tube and emission probe microphone were used for calibration, respectively. The horizontal line represents the commonly used f_2/f_1 of 1.2.

Because the emission level depends on the levels of the stimulus tones, we should predict somewhat different emission growth patterns, depending on which microphone is used for calibration. This trend is illustrated in the data pooled from the eight subjects at 5 kHz (Fig. 7). In Fig. 7(A), the DPOAE level, measured with $L_2=60$ dB SPL and L_1 varied, is generally higher (by typically around 10 dB) when the emission probe microphone (dashed line with open symbols) is used during calibration versus the probe tube microphone (solid line with solid circles). These data exhibit a decline at the highest values of L_1. There is also a large divergence in the DPOAE levels measured with L_1 fixed and L_2 varied [Fig. 7(B)]. The first quarter-wave frequency was near 4 kHz in these subjects. This again illustrates the effect of higher eardrum stimulus levels [for the stimulus frequencies in this figure ($f_1=4167$ Hz, $f_2=5000$ Hz)] when using the emission probe for calibration than when using the probe tube.

C. DPOAE level as a function of stimulus frequency ratio

To determine the effect of varying the stimulus frequency ratio (f_2/f_1) on the DPOAE level, the DPOAE level was measured by fixing f_2 and varying f_1 (yielding frequency ratios of 1.33 to 1.11, in 0.02 increments). The data were obtained from eight adult ears, five males and three females, with ages ranging from 19 to 30 years (mean $=25$ years). Figure 8(A) illustrates the mean data for the level of $2f_1-f_2$ as a function of f_2/f_1 (varied from 1.11 to 1.33) at fixed f_2 frequencies of 2, 4, 8, 10, and 13 kHz, using the probe tube microphone for calibration. A nonmonotonic trend is apparent with varied stimulus frequency ratios for each f_2 frequency. The DPOAE level decreases at narrow and wide ratios. The stimulus frequency ratio that yields the largest DPOAE level decreases as f_2 increases, as reported previously (Harris *et al.*, 1989). The arrows in Fig. 8(A) indicate the primary frequency ratio that yielded the largest DPOAE level for 2- and 10 kHz. The range of f_2/f_1 that evoked measurable emissions decreased with increasing f_2.

The primary frequency ratio producing the largest DPOAE level, for each f_2 frequency, (2 to 13 kHz) was replotted in Fig. 8(B) (solid line with solid symbols). The decreasing trend in this ratio with increasing frequency is clear. The solid line at f_2/f_1 of 1.2 is representative of the stimulus frequency ratio typically used as the default to measure DPOAEs clinically. For comparison purposes, the dashed line with open symbols in Fig. 8(B) represents corresponding data when the emission probe microphone was used as the reference for calibration. A nonmonotonic function is seen with a local increase in the primary frequency ratio producing the largest DPOAE level from 5 to 8 kHz, whereas the general trend was decreasing. The biggest discrepancy between the data collected with the two calibration procedures occurred between 5 and 8 kHz.

D. Group delay

The group delay of the $2f_1-f_2$ DPOAE was calculated from the slope of the distortion-product phase versus distortion-product frequency curve, averaged over a small range of closely spaced distortion-product frequencies. DPOAE group delays were obtained from eight adult ears, the same subjects as just described for estimating the DPOAE level as a function of stimulus frequency ratio. The $2f_1-f_2$ otoacoustic emission group delay was measured by fixing f_2 and varying f_1 (yielding frequency ratios of 1.33 to 1.11, in 0.02 increments). The change in f_1 was small enough so that there were enough points per cycle of the distortion product phase to avoid 2π phase ambiguity. The stimulus levels were 60- and 45 dB SPL for f_1 and f_2, respectively. The frequency of f_2 ranged from 2 to 13 kHz (2, 2.5, 3, 4, 5, 6, 8, 10, and 13 kHz).

Acoustic delays in each of the system components were measured in a small test cavity using a B&K 1/8-in. condenser microphone as a reference. The calibration procedure using EMAV software does not correct the system phase. Thus, the system phase delay was subtracted from the measurements, *post hoc*.

Mean (± 1 SEM) group delays, across subjects, are illustrated in Fig. 9. Group delay measurements made with the probe tube microphone used for calibration are depicted by the solid lines and solid symbols. The DPOAE group delay decreases with increasing frequency, as has been illustrated elsewhere for f_2 frequencies equal to or less than 10 kHz (Kimberley *et al.*, 1993; O Mahoney and Kemp, 1995; Eggermont *et al.*, 1996). Our data show that the pattern of decreasing group delay with increasing frequency extends to at least 13 kHz.

The dashed lines and open symbols in Fig. 9 represent the group delay measurements made using the emission probe microphone for calibration. The pattern of the results is similar to those just reported for the probe tube microphone; however, the group delay measurement using the probe tube microphone for calibration is greater than for the emission probe microphone, especially from 3 to 7 kHz, the range corresponding to the first quarter-wave pressure null at the emission probe location.

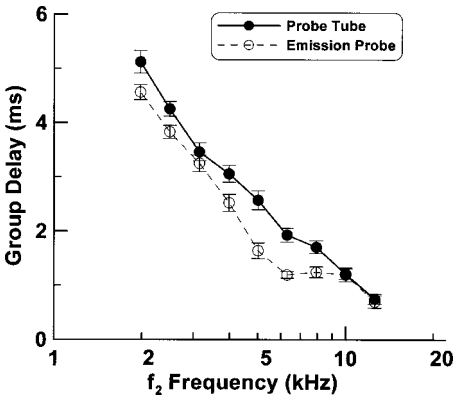


FIG. 9. The group delay, collapsed across subjects, decreases with increasing frequency. Standard error bars are indicated. The solid lines represent data collected with the probe tube microphone for calibration and the dashed lines represent data collected with the emission probe microphone for calibration.

Test–retest DPOAE group delay measurements were made for all subjects. Regression analysis was completed to determine individual repeatability over the two trials. The range of R-squared scores were 0.84 to 0.97 for the probe tube microphone and 0.82 to 0.96 for the emission probe microphone. This result suggested that the DPOAE group delay, made using either calibration procedure, was stable over time within an individual. The R-squared scores were then converted to Z scores (Blalock, 1972), so that a mean R score could be calculated. The mean R scores were 0.97 and 0.98 for the probe tube and emission probe microphones, respectively. Thus, the mean test–retest correlations were very good.

IV. DISCUSSION

We will first discuss our evidence that the high-frequency DPOAEs originate in cochlear mechanics. Then, we will address the influence of differences in the calibration procedures we used.

A. Origin of high-frequency distortion product otoacoustic emissions in humans

These experiments tested the assumption that the origin of DPOAEs in humans elicited with high-frequency stimuli is the same as for DPOAEs measured with low-frequency stimuli. The mechanism of the cochlear amplifier that is thought to generate emissions is still uncertain, which is why we were questioning whether a single mechanism could account for the behavior of otoacoustic emissions over the entire range of hearing. As mentioned earlier, neither the somatic motility (Dallos and Evans, 1995a, b) nor hair-bundle motor (Martin *et al.*, 2000) mechanisms have been proven to be the basis of the cochlear amplifier. Because somatic motility is driven by the ac receptor potential of the outer hair cells, which is low-pass filtered by the membrane time constant, it might be expected that motility contributes to otoacoustic emissions only at lower frequencies (below about 4 kHz). It is plausible that a second mechanism of distortion generation, possibly the hair bundle, is required at higher frequencies. If this is the case, then the two underlying nonlinear functions are likely to be different and characteristics of the distortion generated by the two would also be expected to be different. In this context, it is noteworthy that our findings that DPOAEs measured with varied stimulus parameters demonstrate continuous trends across frequency. These trends support our assumption that both high- and low-frequency DPOAEs have a similar origin. Each of the experiments yielded results indicating that the emissions originate in cochlear mechanics and not in the middle ear. These measurements characterized how the level of the DPOAE depends on the frequencies, levels, and frequency ratios of the stimuli and how the group delay of the emission varied as a function of frequency.

1. Dependence of emission level when stimulus frequencies are varied together with a fixed ratio

DPOAEs were measured with stimuli up to at least 16 kHz in each of the four subjects. The level of the DPOAE was relatively independent of stimulus frequency below 10 kHz. Above 10 kHz, the emission level generally decreased with a pattern that was subject dependent. These patterns suggest a similar origin of DPOAEs over the entire range of frequencies where emissions can be measured. Siegel and Hirohata (1994b) have reported a similar steep decline in DPOAE levels at high frequencies that paralleled the steep cutoff in the subject's audiogram. The cutoff was as low as 10 kHz in some of the older subjects they studied, indicating that the upper limit was imposed by the basal extent of functional organ of Corti and not, for example, determined by the middle ear. In this study, there was not enough data to make meaningful comparisons between the behavioral thresholds and the DPOAE levels measured for these four subjects.

2. Dependence of emission level with varied levels of the stimuli

Studies have demonstrated different DPOAE level I/O functions for different primary frequencies when using a fixed level difference between the primary frequencies. Using a fixed 15-dB primary level difference, Harris and Probst (1990) reported more linear DPOAE functions as the primary frequencies were increased. Stover, Gorga *et al.* (1996) reported varied overall shapes of DPOAE I/O functions across f_2 frequency with a fixed 10-dB stimulus level difference. Specifically, some of the I/O functions showed nonmonotonicities, evidence of saturation, or continual growth even for the highest stimulus levels. Stover and Norton (1993) and Nelson and Kimberley (1992) also reported similar variations for I/O patterns. The results reported by Stover, Gorga *et al.* (1996) are in agreement with those of Robles *et al.* (1997), who found that the $2f_1-f_2$ magnitude saturated at moderately high stimulus intensities in the basilar membrane of chinchilla, for equal level primary tones (0-dB primary level difference) and for differing primary frequency separations.

In agreement with Harris and Probst (1990) and Popelka *et al.* (1993), when incrementing both stimulus levels ($L_1 > L_2$ by 15 dB SPL or $L_1 = L_2$, respectively), we have demonstrated a monotonic increase in the level of the DPOAE of approximately 1 dB/dB, from the point at which the DPOAE level leaves the noise floor. This finding is similar to what

has been reported in psychoacoustic measurements of the $2f_1-f_2$ cancellation level with equal-level primary tones (Shannon and Houtgast, 1980). The rate of growth of the DPOAE at low levels was approximately 1 dB/dB at all frequencies, as reported in otoacoustic emission studies (reviewed by Nelson and Zhou, 1996). The monotonic behavior of the level functions reported here may be due to our use of relatively low-level stimuli. Calibration procedures may have also played a role, because we found fewer examples of nonmonotonicity when the stimulus levels were calibrated using the probe tube at the eardrum (see the later discussion).

The results of DPOAEs measured with one stimulus level fixed and the other stimulus level varied are similar to those measured psychoacoustically (Goldstein, 1967; Smoorenburg, 1972; Shannon and Houtgast, 1980; Zwicker, 1981) and physiologically at the level of cochlear mechanics (Rhode and Cooper, 1993; Robles *et al.*, 1997; Cooper and Rhode, 1997). As in these other studies, the DPOAE level has been shown to be more dependent on L_1 than L_2 in rabbits, guinea pigs, and humans (Whitehead *et al.*, 1992; Brown and Gaskill, 1990; Gaskill and Brown, 1990; Whitehead, Stagner *et al.*, 1995). In this study, when incrementing one stimulus level relative to the other, the dependence of the level of the $2f_1-f_2$ otoacoustic emission on the levels of the stimuli is similar for f_2 frequencies from 2 to 16 kHz (Fig. 6). As reported by others, the DPOAE level is more strongly dependent on L_1 than L_2. The only exception in our data is the case where L_1 was fixed at 70 dB SPL.

The maximum level of the DPOAE is measured with $L_1>L_2$ for all stimulus frequencies and levels, at which a maximum could be defined in our data. This result is similar to DPOAE findings at lower frequencies (Gaskill and Brown, 1990), cancellation levels of the combination tone $2f_1-f_2$ measured psychoacoustically (Zwicker, 1981), and for intermodulation distortion products measured from the basilar membrane (Robles *et al.*, 1997). Additionally, our data are consistent with the results of other studies indicating that the stimulus level difference yielding the largest DPOAE level decreases with increasing level of the fixed stimulus.

These similarities in the stimulus-level dependence of the DPOAE level across stimulus frequency further support our suggestion that the nonlinear cochlear mechanisms, presumably biological, which generate emissions are similar in all regions of the cochlea. We presume that the DPOAE level would have increased with increasing L_2 at lower fixed levels of L_1 had we varied L_2 below 30 dB SPL, but we have not studied this range of stimulus levels (Dorn *et al.*, 2001).

3. Dependence of emission level with varied primary frequency ratios

A prominent argument that psychophysical combination tones were generated in the cochlea rather than the middle ear cited evidence for a steep drop in the combination tone levels as the ratio of the stimulus frequencies (f_2/f_1) was increased (Zwicker, 1981). It was argued that this trend resulted from decreased overlap of the excitation patterns of the stimuli. The level of DPOAEs has also been shown to vary depending on the frequency ratio of the primary stimuli. This trend has been shown in studies using rabbits (Lonsbury-Martin *et al.*, 1987; Whitehead *et al.*, 1990, 1992), guinea pigs (Brown and Gaskill, 1990), and humans (Wilson, 1980; Harris *et al.*, 1989; Gaskill and Brown, 1990; Abdala, 1996; Stover *et al.*, 1999). Not only does the DPOAE level decrease as a function of the stimulus frequency ratio for large ratios (i.e., $f_2/f_1>1.2$), but the DPOAE level also decreases at narrow ratios. This nonmonotonic trend has been shown using both f_1 (Abdala, 1996; Stover *et al.*, 1999) and f_2 (Gaskill and Brown, 1990) sweep paradigms. Harris *et al.* (1989) and Abdala (1996) both reported that the slopes of the declines at both the narrow and wide stimulus frequency ratios were steeper for higher frequencies. Additionally, the slopes tended to be steeper on the wide ratio side of the nonmonotonic function. In agreement with these previous studies, a nonmonotonic function of f_2/f_1 was found in this study for f_2 frequencies from 2 to 13 kHz with varied stimulus frequency ratios. The fact that large DPOAEs were measured over a more restricted range of f_2/f_1 at high frequencies might suggest sharper mechanical tuning [Fig. 8(A)]. However, because the slopes of these functions may be intensity dependent and the efficiency of middle-ear transmission is frequency dependent, we cannot exclude a potentially confounding level dependence.

DPOAE results with varied stimulus frequency ratios do not agree with all psychoacoustic and physiologic data. With varied stimulus frequency ratios, Cooper and Rhode (1997) demonstrated that the basilar-membrane response of guinea pigs measured with $2f_1-f_2$ near the characteristic frequency resulted in a nonmonotonic function, as has been reported for DPOAEs. However, a decline in the levels of intermodulation distortion products at narrow stimulus frequency ratios has not been reported for some basilar-membrane measurements (Robles *et al.*, 1997) or for most studies of the psychoacoustical combination tone $2f_1-f_2$ (Plomp, 1965; Smoorenburg, 1972; Shannon and Houtgast, 1980; Wilson, 1980). Despite the discrepancy between these various studies for narrow ratios, the steep decline in emission levels for wide ratios strongly indicates an origin in cochlear mechanics rather than the middle ear.

Studies of DPOAEs have reported that as the primary frequency increased, the stimulus frequency ratio yielding the largest DPOAE level decreased in rabbits (Lonsbury-Martin *et al.*, 1987; Whitehead *et al.*, 1990, 1992), bats (Kössl, 1994; Frank and Kössl, 1995), and in humans (Harris *et al.*, 1989; Abdala, 1996; Bowman *et al.*, 2000). Typically, in the current study, the stimulus frequency ratio generating the largest DPOAE level decreased with increasing frequency [Fig. 8(B)]. In our data, an f_2/f_1 of 1.2 did not generate the largest DPOAE level for all primary frequencies tested, as is typically used in clinical settings. For example, DPOAE levels could be as much as 5 and 10 dB SPL greater for 13 and 10 kHz, respectively, with f_2/f_1 ratios of approximately 1.15 versus 1.2. Our findings agree with those of Bowman *et al.* (2000), who reported that a stimulus frequency ratio of 1.16 generated the largest DPOAE level at 9 kHz versus 1.25 for 1 kHz, in humans. Physiological measurements have provided evidence for sharper mechanical tuning with increasing center frequency (e.g., Evans, 1972,

Kiang and Moxon, 1974, Dallas and Harris, 1978, Patuzzi and Robertson, 1988). Direct mechanical studies reveal relatively broader tuning in the apex of the chinchilla cochlea versus sharper tuning in the base, with the sharpest tuning occurring in the hook region of the guinea pig cochlea (Cooper and Rhode, 1997). Thus, the fact that the f_2/f_1 generating the largest DPOAE level changes with frequency indicates a cochlear origin of DPOAEs.

4. DPOAE group delay

Estimating the group delay of the DPOAE $2f_1-f_2$ is usually measured by fixing f_2 and varying f_1 (f_1 sweep) and then calculating the slope of the distortion-product phase versus distortion-product frequency curve. Group delay can also be measured by fixing f_1 and varying f_2 (f_2 sweep) or by varying the primaries together to yield a constant frequency ratio (iso-f_2/f_1). Interpreting DPOAE group delay is more difficult than for a single continuous tone (O Mahoney and Kemp, 1995) and cannot be related directly to cochlear filtering (Shera *et al.*, 2000). In these experiments we present group delay data without interpretation, simply to compare this behavior of the emissions over the range of stimulus frequencies studied.

It has been demonstrated, using primary frequency stimuli equal to or less than 10 kHz, that group delay increased for $2f_1-f_2$ as the primary frequencies were decreased, commonly assumed to reflect a progressively more apical origin of the distortion product in the cochlea. This result was true using either an f_1 sweep paradigm (Kimberley *et al.*, 1993; Whitehead *et al.*, 1994; Stover *et al.*, 1994; Stover, Neely, and Gorga, 1996; Eggermont *et al.*, 1996; Moulin and Kemp, 1996; Whitehead *et al.*, 1996; Shera *et al.*, 2000) or an f_2 sweep paradigm (Whitehead *et al.*, 1994; O Mahoney and Kemp, 1995; Moulin and Kemp, 1996; Whitehead *et al.*, 1996; Shera *et al.*, 2000).

Cochlear mechanics appears to make the major contribution to the measured group delay for low-frequency stimuli. The group delay in the ear canal and middle ear has usually been assumed negligible. For example, the propagation delay in the ear canal from the emission probe to the eardrum is around 0.06 ms for most frequencies (O Mahoney and Kemp, 1995). However, we have demonstrated that this assumption is not always justified (Dreisbach and Siegel, 1995). In particular, we found that standing waves can contribute directly to the group delay measure for frequencies near quarter-wave pressure minima at the emission probe, and also indirectly by influencing the stimulus levels at the eardrum.

The group delay of the $2f_1-f_2$ distortion product emission measured in this study is generally consistent with previous reports (Kemp and Brown, 1983; Kimberley *et al.*, 1993; Stover *et al.*, 1994; Eggermont *et al.*, 1996) for stimulus frequencies up to 10 kHz. Specifically, Kimberley *et al.*, (1993) obtained their data with the same stimulus levels and frequency ratios of the stimulus tones that we used in this study. The group delay in this study continued to decline at 13 kHz (Fig. 9) to approximately 0.75 ms. This observation further supports the interpretation that the DPOAEs measured for higher stimulus frequencies are generated in the most basal region of the cochlea. The previous report that emissions are coexistent with the range of high-frequency hearing in individual subjects (Siegel and Hirohata, 1994b) is consistent with this view. Additionally, the group delay was very similar for two different trial periods. It would be interesting to examine whether group delay measures can be used to detect changes in cochlear function over time.

Measurements in this study can be compared to more invasive measurements made at the level of the basilar membrane. For a characteristic frequency (CF) of 10 kHz, a single-tone group delay of 1.0 ms was reported in chinchilla (Ruggero *et al.*, 1997). In this study, the DPOAE group delay at 10 kHz, measured with the fixed f_2 paradigm, was 1.21 ms, similar to the value measured in chinchillas using a single tone. The group delay at CF decreased at 11 kHz to 0.56 ms in chinchilla and we report a decreased DPOAE group delay of 0.75 ms at 13 kHz. Although there is reason to doubt a direct correspondence, the trends of decreasing group delay with increasing frequency are similar for measurements made at the level of the basilar membrane in chinchilla and measurements made using DPOAEs in humans.

B. Calibration

The system used in this study, the ER10S microphone with an endoscope, was capable of measuring distortion-product emissions at high frequencies (>8 kHz) and allowed a probe tube microphone to be placed near the eardrum under optical control. The probe tube placed at the eardrum, for calibration purposes, allowed better control over the stimulus levels presented versus a more traditional setup with the calibration completed at the emission probe. This system accurately controlled eardrum sound-pressure levels across a wider range of frequencies than does conventional in-the-ear calibration using the emission probe, and this minimizes the effect of the acoustics of the ear canal and simplifies the trends in the emission data. Effects of calibration errors were clearly demonstrated in each of the data collection paradigms used to characterize the emissions.

1. Influence of calibration procedure on the dependence of the emission level when stimulus frequencies are varied together with a fixed ratio

Generally, no matter which microphone is used for calibration, patterns of the DPOAE level with frequency are similar. However, a larger emission level often was measured when the emission probe, versus the probe tube, microphone was used for calibration. For the four subjects shown in Fig. 4, the emission level differences were greatest from 2.2 to 4.3 kHz and from 6.6 to 20 kHz, the ranges of stimulus frequencies for which quarter-wave pressure nulls were present in the emission probe microphone response for these subjects [see Fig. 2(A)]. The differences in emission levels between the two calibration procedures were as great as 20 dB. The higher emission levels for stimuli near the quarter-wave frequencies result from local increases in the stimulus levels at the eardrum (Siegel, 1994, 2002). With the probe tube calibration, the emission level is relatively independent of stimulus frequency from 1 kHz to the upper limit of the subject's normal hearing range (Siegel and Hirohata, 1994b).

2. Influence of calibration procedure on the dependence of emission level on stimulus levels

Because the emission level depends on the levels of the stimulus tones, we should predict somewhat different emission growth patterns, depending on which microphone is used for calibration. The emission probe and probe tube microphone calibrations yielded similar growth curves of the DPOAE level with increased stimulus levels (both varied together) at 2 and 13 kHz, but differed at 4, 6, 8, and 10 kHz. At these frequencies, the distortion-product level was greater for the emission probe than for the probe tube calibration procedure. This is the frequency region most affected by standing waves. We have demonstrated that using either calibration procedure (emission probe versus probe tube) produces a monotonic increase in the level of the DPOAE, from the point at which the DPOAE level leaves the noise floor, as the stimulus levels are increased simultaneously. However, a discrepancy between the measured DPOAE levels exists between the two calibration procedures in the frequency region of 4 to 10 kHz. This is most likely due to the fact that for frequencies where pressure nulls due to standing waves are present at the emission probe, the stimulus level at the eardrum is greater when the emission probe microphone is used as the reference for calibration. Additionally, when the emission probe microphone was used during calibration, saturation of the functions, especially at 6 kHz, were noted, as reported by Stover *et al.* (1996) in humans and Robles *et al.* (1997) in chinchilla. We suspect that saturation would have been more evident in our data with the probe tube calibration had we used higher stimulus levels. The lack of a difference between the two calibration procedures at 13 kHz could be because pressure nulls at frequencies greater than 10 kHz are sharper than at lower frequencies and vary significantly between subjects. The data reported were averaged across subjects. Thus, individual variations between the two calibration procedures may be diminished at 13 kHz.

Not surprisingly, the calibration procedure also influenced the level dependence of emissions when the level of one tone was varied. Again, the greatest effect of the calibration procedure was seen at 5 kHz. With the emission probe used for calibration for an f_2 of 5 kHz, the eardrum stimulus levels are higher than the intended values by an amount determined by the proximity of each stimulus component to the first quarter-wave frequency. The DPOAE level, measured with L_2 equal to 60 dB SPL and L_1 varied, is generally higher (by typically around 10 dB) when the emission probe microphone is used during calibration [Fig. 7(A)]. These data exhibit a decline at the highest values of L_1. There is also a large divergence in the DPOAE levels measured with L_1 fixed and L_2 varied [Fig. 7(B)]. The first quarter-wave frequency was near 4 kHz in our subjects. Correspondingly, the stimulus condition most affected by the calibration procedures was for stimulus $f_1 = 4167$ Hz and $f_2 = 5000$ Hz.

The dependence of the DPOAE on the levels of the stimulus tones is therefore substantially influenced by artifacts in calibrating stimulus levels resulting from the presence of standing waves in the ear canal.

3. Dependence of emission level with varied primary frequency ratios

When comparing the results from the different calibration methods, the stimulus frequency ratio yielding the largest DPOAE level for the probe tube calibration generally decreased with increasing f_2, whereas a nonmonotonic function with a local peak in the 5- to 8-kHz frequency region was seen when the emission probe microphone was used during calibration.

The discrepancy in the f_2/f_1 that yielded the peak DPOAE between the emission probe and probe tube microphone calibration in the frequency region from 5 to 8 kHz is probably due to the relatively poor control over eardrum stimulus levels when the emission probe microphone is used as the reference for calibration. Wilson (1980), Harris *et al.* (1989), and Whitehead *et al.* (1990, 1992) reported that the f_2/f_1 generating the peak DPOAE depends on stimulus level, increasing with increasing stimulus levels. Thus, our data showing a local maximum from 5 to 8 kHz, collected using the emission probe for calibration, are likely to result from increased eardrum stimulus levels in this frequency range.

4. DPOAE group delay

Comparisons were made to determine if the method of calibration (probe tube versus emission probe microphone) had an effect on the DPOAE group delay. Group delay decreased with increasing frequency for each calibration method used. However, the group delay measurement for the probe tube calibration (Fig. 9) was greater than for the emission probe microphone, especially from 3 to 7 kHz, the area most affected by the quarter-wave pressure nulls. These differences between the group delay measurements for the two calibration methods are likely to result from the presence of standing waves in the ear canal. The first possible mechanism is the increase in the ear-canal group delay near the quarter-wave frequencies, which could affect the phase of the f_1 stimulus at the eardrum, and thus the measured group delay of the emission. Because the $2f_1 - f_2$ emission also propagates in an ear canal, its phase may also be affected by standing waves, but this effect on the outward propagation of emission signals has not been characterized.

The second possible contribution of standing waves to the group delay measure is the difference in stimulus level at the eardrum near the quarter-wave frequencies that is the inevitable consequence of the two calibration methods. Specifically, the stimulus level at the eardrum should be higher when the emission probe microphone is used during calibration. It has been reported that the DPOAE group delay decreases with increasing stimulus level (Kimberley *et al.*, 1993; Stover *et al.*, 1994; Bowman *et al.*, 1997). The first quarter-wave peak in the ear-canal transfer functions of our subjects ranged from 4 to 7 kHz. It is therefore expected that the greatest difference in the emission group delay for the two calibration methods should be when the stimuli are in this frequency range. Calibration with the emission probe should yield the shorter delays, especially in this frequency region. This trend is clearly demonstrated in our subject population and in others (O Mahoney and Kemp, 1995;

Moulin and Kemp, 1996). At 4 kHz, O Mahoney and Kemp (1995) found a significant difference between group delay measurements made using a ''single-fit'' versus ''multiple-fit'' method of probe insertion. Also, a large intersubject range of latency values was obtained. Their postulated explanation for this finding was the existence of multiple sources of the DPOAE. Similarly, Moulin and Kemp (1996) reported that among the general pattern of decreasing latency as frequency increased, a nonmonotonic behavior was observed, which occurred in the 4-to 5-kHz region for both f_2 and f_1 sweeps. From our results in this study, it seems likely that the discrepancies seen in these studies in the 4-kHz region were actually due to the first quarter-wave pressure null in the ear canals of their subjects. The greatest difference in group delay between the emission probe and probe tube calibrations in our pooled data is around 0.6 ms and is highly significant ($p<0.001$). The fact that there was little discrepancy in group delay for higher frequencies is consistent with the observation that the level dependence of the DPOAE group delay decreases with increasing frequency (Kimberley *et al.*, 1993).

V. SUMMARY

Given that the results with varied stimulus frequencies, levels, ratios, and group delay measures of the DPOAE elicited with high-frequency stimuli are similar to those measured with low-frequency stimuli, it can be concluded that the same biological process in the inner ear responsible for low-frequency DPOAEs is responsible for high-frequency DPOAEs. Additionally, because DPOAEs were measured with stimuli greater than 8 kHz, comparisons can be made with intermodulation distortion products measured at the level of the basilar membrane.

Consistent with a previous report (Siegel and Hirohata, 1994a) the present results demonstrate that calibration artifacts influence all the emission behaviors we studied. The frequency range most affected by these artifacts was from 3 to 7 kHz, corresponding to the frequency range where the first quarter-wave pressure null appears at the emission probe inlet. These artifacts can produce differences in the levels of emissions as large as 20 dB, but also influence other measures such as the group delay and the dependence of the emission level on the frequency ratio of the stimuli. For frequencies greater than 10 kHz, there is no consistent discrepancy between the data obtained with the two different calibration procedures. This result is consistent with other studies in suggesting that probe tube measurements do not adequately define the input level to the human ear at frequencies greater than 10 kHz (Stinson, 1985).

Using a consistent definition for the input level to the ear (i.e., eardrum sound-pressure level), as was done in this study, might enhance the consistency of emission levels across frequency when compared with the conventional DPOAE calibration procedure using the emission probe in the subject's ear. When the probe tube was used for calibration, our data revealed that the DPOAE level is relatively insensitive to L_2 as long as $L_1>L_2$ at the eardrum. This result is consistent with published reports demonstrating shallower growth of the DPOAE level with increasing L_2 versus increasing L_1, using in-the-ear calibrations (Gaskill and Brown, 1990; Hauser and Probst, 1991; Whitehead, McCoy *et al.*, 1995). So, even if the traditional in-the-ear calibration using the emission probe is used, setting L_2 20 to 30 dB below L_1 should make DPOAE measurements relatively insensitive to variations in L_2 due to calibration errors for f_2 frequencies near the quarter-wave pressure nulls. This would more or less guarantee that L_1 would be greater than L_2 at the eardrum, placing L_2 in a range where variations would have little effect on the emission level. However, this does not remove the influence on emission levels due to calibration artifacts affecting the level of f_1 resulting from standing waves. But, by reducing variability due to variations in L_2, the reliability of DPOAE measurements may be improved in clinical applications where it is inconvenient to place a probe tube near the eardrum with the emission probe in place. Other alternatives to in-the-ear calibration should be explored, including calibration of the emission probe in a coupler.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. R01 DC02021. We are grateful to Etymotic Research and Electro FiberOptics Corp. for building the customized devices. We would also like to thank two anonymous reviewers for their helpful suggestions on an earlier version of this manuscript.

Abdala, C. (**1996**). ''Distortion product otoacoustic emission ($2f_1-f_2$) amplitude as a function of f_2/f_1 frequency ratio and primary tone level separation in human adults and neonates,'' J. Acoust. Soc. Am. **100**, 3726–3740.

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

Blalock, H. M. (**1972**). *Social Statistics*, 2nd ed. (McGraw-Hill, New York).

Bowman, D. M., Brown, D. K., Eggermont, J. J., and Kimberley, B. P. (**1997**). ''The effect of sound intensity on f_1-sweep and f_2-sweep distortion product otoacoustic emissions phase delay estimates in human adults,'' J. Acoust. Soc. Am. **101**, 1550–1559.

Bowman, D. M., Brown, D. K., and Kimberley, B. P. (**2000**). ''An examination of gender differences in DPOAE phase delay measurements in normal-hearing adults,'' Hear. Res. **142**, 1–11.

Brass, D., and Locke, A. (**1997**). ''The effect of the evanescent wave upon acoustic measurements in the human ear canal,'' J. Acoust. Soc. Am. **101**, 2164–2175.

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 Gaskill, S. A. (**1990**). ''Measurement of acoustic distortion reveals underlying similarities between human and rodent mechanical responses,'' J. Acoust. Soc. Am. **88**, 840–849.

Brown, A. M., McDowell, B., and Forge, A. (**1989**). ''Effects of chronic gentamicin treatment on hair cells can be monitored using acoustic distortion products,'' Hear. Res. **42**, 143–156.

Chan, C. K., Musicant, A. D., and Hind, J. E. (**1993**). ''An insert earphone system for delivery of spectrally shaped signals for physiological studies,'' J. Acoust. Soc. Am. **93**, 1496–1501.

Cooper, N. P., and Rhode, W. S. (**1997**). ''Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea,'' J. Neurophysiol. **78**, 261–270.

Dallos, P., and Evans, B. N. (**1995a**). ''High-frequency motility of outer hair cells and the cochlear amplifier,'' Science **267**, 2006–2009.

Dallos, P., and Evans, B. N. (**1995b**). ''High-frequency outer hair cell motility: Corrections and addendum,'' Science **268**, 1420–1421.

Dallos, P., and Harris, D. (**1978**). ''Properties of auditory nerve responses in absence of outer hair cells,'' J. Neurophysiol. **41**, 365–383.

Dorn, P. A., Konrad-Martin, D., Neely, S. T., Keefe, D. H., Cyr, E., and Gorga, M. P. (**2001**). ''DPOAE input/output functions in normal and impaired human ears,'' Assoc. Res. Otolar. Abstracts **24**, 8.

Dreisbach, L. E., and Siegel, J. H. (**1995**). ''The effect of ear canal acoustics on high frequency distortion-product otoacoustic emission group delay measurements,'' International Evoked Response Audiometry Study Group Abstracts **14**, 17.

Eggermont, J. J., Brown, D. K., Ponton, C. W., and Kimberley, B. P. (**1996**). ''Comparison of distortion product otoacoustic emission (DPOAE) and auditory brain stem response (ABR) traveling wave delay measurements suggests frequency-specific synapse maturation,'' Ear Hear. **17**, 386–394.

Evans, E. F. (**1972**). ''The frequency response and other properties of single fibres in the guinea-pig cochlear nerve.'' J. Physiology **226**, 263–87.

Frank, T., and Dreisbach, L. E. (**1991**). ''Repeatability of high-frequency thresholds,'' Ear Hear. **12**, 294–295.

Frank, G., and Kössl, M. (**1995**). ''The shape of 2 f_1-f_2 suppression tuning curves reflects basilar membrane specializations in the mustached bat, *Pteronotus parnellii*,'' Hear. Res. **83**, 151–160.

Gaskill, S. A., and Brown, A. M. (**1990**). ''The behavior of the acoustic distortion product, 2 $f1-f2$, from the human ear and its relation to auditory sensitivity,'' J. Acoust. Soc. Am. **88**, 821–839.

Goldstein, J. L. (**1967**). ''Auditory nonlinearity,'' J. Acoust. Soc. Am. **41**, 676–689.

Harris, F. P., Lonsbury-Martin, B. L., Stagner, B. B., Coats, A. C., and Martin, G. K. (**1989**). ''Acoustic distortion products in humans: systematic changes in amplitudes as a function of $f2/f1$ ratio,'' J. Acoust. Soc. Am. **85**, 220–229.

Harris, F. P., and Probst, R. (**1990**). ''Growth functions of tone burst evoked and distortion-product otoacoustic emissions in humans,'' in *Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer, New York), pp. 178–185.

Hauser, R., and Probst, R. (**1991**). ''The influence of systematic primary-tone level variation L_2-L_1 on the acoustic distortion product emission 2 f_1-f_2 in normal human ears,'' J. Acoust. Soc. Am. **89**, 280–286.

Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T., and Brown, A. M. (**1983**). ''An integrated view of cochlear mechanical nonlinearities observable from the ear canal'' in *Mechanics of Hearing*, edited by E. de Boer and M. A. Viergever (Delft University Press, Delft, The Netherlands), pp. 75–82.

Khanna, S. M., and Leonard, D. G. B. (**1986**). ''Relationship between basilar membrane tuning and hair cell condition,'' Hear. Res. **23**, 55–70.

Kiang, N. Y., and Moxon, E. C. (**1974**). ''Tails of tuning curves of auditory-nerve fibers.'' J. Acoust. Soc. Am. **55**, 620–630.

Kim, D. O. (**1980**). ''Cochlear mechanics: Implications of electrophysiological and acoustical observations,'' Hear. Res. **2**, 297–317.

Kim, D. O., Molnar, C. E., and Matthews, J. W. (**1980**). ''Cochlear mechanics: Nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal sound pressure,'' J. Acoust. Soc. Am. **67**, 1704–1721.

Kimberley, B. P., Brown, D. K., and Eggermont, J. J. (**1993**). ''Measuring human cochlear traveling wave delay using distortion product emission phase responses,'' J. Acoust. Soc. Am. **94**, 1343–1350.

Kössl, M. (**1994**). ''Otoacoustic emissions from the cochlea of the 'constant frequency' bats, *Pteronotus parnelli and Rhinolophus rouxi*,'' Hear. Res. **72**, 59–72.

Liberman, M. C., and Dodds, L. W. (**1984**). ''Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves,'' Hear. Res. **16**, 55–74.

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**, 173–189.

Martin, P., Mehta, A. D., and Hudspeth, A. J. (**2000**). ''Negative hair-bundle stiffness betrays a mechanism for mechanical amplification by the hair cell,'' Proc. Natl. Acad. Sci. U.S.A. **97**, 12026–12031.

Moulin, A., and Kemp, D. T. (**1996**). ''Multicomponent acoustic distortion product otoacoustic emission phase in humans. I. General characteristics,'' J. Acoust. Soc. Am. **100**, 1617–1639.

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

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.

Nelson, D. A., and Zhou, J. Z. (**1996**). ''Slopes of distortion-product otoacoustic emission growth curves corrected for noise-floor levels,'' J. Acoust. Soc. Am. **99**, 468–474.

O Mahoney, C. F., and Kemp, D. T. (**1995**). ''Distortion product otoacoustic emission delay measurement in human ears,'' J. Acoust. Soc. Am. **97**, 3721–3735.

Ohlemiller, K. K., and Siegel, J. H. (**1994**). ''Cochlear basal and apical differences reflected in the effects of cooling on responses of single auditory nerve fibers,'' Hear. Res. **80**, 174–190.

Patuzzi, R., and Robertson, D. (**1988**). ''Tuning in the mammalian cochlea,'' Physiological Review **68**, 1009–1082.

Plomp, R. (**1965**). ''Detectability threshold for combination tones,'' J. Acoust. Soc. Am. **37**, 1110–1123.

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.

Rhode, W. S., and Cooper, N. P. (**1993**). ''Two-tone suppression and distortion production on the basilar membrane in the hook region of cat and guinea pig cochleae,'' Hear. Res. **66**, 31–45.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1997**). ''Two-tone distortion on the basilar membrane of the chinchilla cochlea,'' J. Neurophysiol. **77**, 2385–2399.

Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neurosci. **11**, 1057–1067.

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.

Shannon, R. V., and Houtgast, T. (**1980**). ''Psychophysical measurements relating suppression and combination tones,'' J. Acoust. Soc. Am. **68**, 825–829.

Shera, C. A., Talmadge, C. L., and Tubis, A. (**2000**). ''Interrelations among distortion-product phase-gradient delays: Their connection to scaling symmetry and its breaking,'' J. Acoust. Soc. Am. **108**, 2933–2948.

Siegel, J. H. (**2002**). ''Calibrating otoacoustic emission probes,'' in *Otoacoustic Emissions: Clinical Applications*, 2nd ed., edited by M. S. Robinette and T. J. Glattke (Thieme, New York), pp. 416–441.

Siegel, J. H. (**1994**). ''Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes,'' J. Acoust. Soc. Am. **95**, 2589–2597.

Siegel, J. H., and Dreisbach, L. E. (**1995**). ''Optical placement of a probe tube in the occluded human ear canal,'' Assoc. Res. Otolar. Abstracts **18**, 119.

Siegel, J. H., and Hirohata, E. T. (**1994a**). ''Sound calibration and distortion product otoacoustic emissions at high frequencies,'' Hear. Res. **80**, 146–152.

Siegel, J. H., and Hirohata, E. T. (**1994b**). ''Ultra-high-frequency distortion-product emissions and thresholds in humans,'' Assoc. Res. Otolar. Abstracts **17**, 55.

Smoorenburg, G. F. (**1972**). ''Combination tones and their origin,'' J. Acoust. Soc. Am. **52**, 615–632.

Stinson, M. R. (**1985**). ''The spatial distribution of sound pressure within scaled replicas of the human ear canal,'' J. Acoust. Soc. Am. **78**, 1596–1602.

Stover, L. J., Gorga, M. P., Neely, S. T., and Montoya, D. (**1996**). ''Toward optimizing the clinical utility of distortion product otoacoustic emission measurements,'' J. Acoust. Soc. Am. **100**, 956–967.

Stover, L. J., Neely, S. T., and Gorga, M. P. (**1994**). ''Explorations of the 'filter' of distortion product otoacoustic emissions,'' Assoc. Res. Otolar. Abstracts **17**, 54.

Stover, L. J., Neely, S. T., and Gorga, M. P. (**1996**). ''Latency and multiple sources of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **99**, 1016–1024.

Stover, L. J., Neely, S. T., and Gorga, M. P. (**1999**). ''Cochlear generation of intermodulation distortion revealed by DPOAE frequency functions in normal and impaired ears,'' J. Acoust. Soc. Am. **106**, 2669–2678.

Stover, L. J., and Norton, S. J. (**1993**). ''The effects of aging on otoacoustic emissions,'' J. Acoust. Soc. Am. **94**, 2670–2681.

Whitehead, M. L., Lonsbury-Martin, B. L., and Martin, G. K. (**1990**). ''Actively and passively generated acoustic distortion at 2 f_1-f_2 in rabbits,'' in *Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer, New York), pp. 243–250.

Whitehead, M. L., Lonsbury-Martin, B. L., and Martin, G. K. (**1992**). ''Evi-

dence for two discrete sources of $2f_1-f_2$ distortion-product otoacoustic emission in rabbit. I. Differential dependence on stimulus parameters,'' J. Acoust. Soc. Am. **91**, 1587–1607.

Whitehead, M. L., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). ''Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. I. Effects of decreasing L_2 below L_1,'' J. Acoust. Soc. Am. **97**, 2346–2358.

Whitehead, M. L., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**1994**). ''Comparison of the onset latency and group latency of distortion-product otoacoustic emissions in human ears,'' Assoc. Res. Otolar. Abstracts **17**, 46.

Whitehead, M. L., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). ''Effects of ear-canal standing waves on measurements of distortion-product otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 3200–3214.

Whitehead, M. L., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (**1996**). ''Visualization of the onset of distortion-product otoacoustic emissions, and measurement of their latency,'' J. Acoust. Soc. Am. **100**, 1663–1679.

Wiederhold, M. L., Mahoney, J. W., and Kellogg, D. L. (**1986**). ''Acoustic overstimulation reduces $2f_1-f_2$ cochlear emissions at all levels in the cat,'' in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer, New York), pp. 322–329.

Wilson, J. P. (**1980**). ''The combination tone, $2f_1-f_2$, in psychophysics and ear-canal recording,'' in *Psychophysical, Physiological and Behavioral Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft University Press, Delft, The Netherlands), pp. 43–52.

Zwicker, E. (**1981**). ''Dependence of level and phase of the $(2f_1-f_2)$-cancellation tone on frequency range, frequency difference, level of primaries, and subject,'' J. Acoust. Soc. Am. **70**, 1277–1288.

Second-order modulation detection thresholds for pure-tone and narrow-band noise carriers

C. Lorenzi
Laboratoire de Psychologie Expérimentale, UMR CNRS 8581, Institut de Psychologie, Université René Descartes Paris V, 71, Av. Edouard Vaillant, 92774 Boulogne-Billancourt, France

M. I. G. Simpson, R. E. Millman, T. D. Griffiths, W. P. Woods, A. Rees, and G. G. R. Green
Department of Physiological Sciences, The Medical School, Framlington Place, Newcastle-Upon-Tyne NE2 4HH, United Kingdom

(Received 1 November 2000; revised 29 March 2001; accepted 23 July 2001)

Modulation perception has typically been characterized by measuring detection thresholds for sinusoidally amplitude-modulated (SAM) signals. This study uses multicomponent modulations. ''Second-order'' temporal modulation transfer functions (TMTFs) measure detection thresholds for a sinusoidal modulation of the modulation waveform of a SAM signal [Lorenzi *et al.*, J. Acoust. Soc. Am. **110**, 1030–2038 (2001)]. The SAM signal therefore acts as a ''carrier'' stimulus of frequency f_m, and sinusoidal modulation of the SAM signal's modulation depth (at rate f'_m) generates two additional components in the modulation spectrum at $f_m - f'_m$ and $f_m + f'_m$. There is no spectral energy at the envelope beat frequency f'_m in the modulation spectrum of the ''physical'' stimulus. In the present study, second-order TMTFs were measured for three listeners when f_m was 16, 64, and 256 Hz. The carrier was either a 5-kHz pure tone or a narrow-band noise with center frequency and bandwidth of 5 kHz and 2 Hz, respectively. The narrow-band noise carrier was used to prevent listeners from detecting spectral energy at the beat frequency f'_m in the ''internal'' stimuli's modulation spectrum. The results show that, for the 5-kHz pure-tone carrier, second-order TMTFs are nearly low pass in shape; the overall sensitivity and cutoff frequency measured on these second-order TMTFs increase when f_m increases from 16 to 256 Hz. For the 2-Hz-wide narrow-band noise carrier, second-order TMTFs are nearly flat in shape for $f_m = 16$ and 64 Hz, and they show a high-pass segment for $f_m = 256$ Hz. These results suggest that detection of spectral energy at the envelope beat frequency contributes in part to the detection of second-order modulation. This is consistent with the idea that nonlinear mechanisms in the auditory pathway produce an audible distortion component at the envelope beat frequency in the internal modulation spectrum of the sounds.
[DOI: 10.1121/1.1406160]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Mk [SPB]

I. INTRODUCTION

This paper addresses the mechanisms that allow human listeners to follow changes in the temporal envelope of sounds. A general approach to describe amplitude modulation resolution is to measure detection thresholds for sinusoidal amplitude modulation (SAM) as a function of modulation frequency (Viemeister, 1979). The resulting temporal modulation transfer function (TMTF) relates modulation detection threshold to modulation frequency. When measured with a broadband noise carrier, TMTFs show a typical low-pass shape. The linear envelope detector model (e.g., Viemeister, 1979; Forrest and Green, 1987; Strickland and Viemeister, 1996) accounts for the low-pass shape of TMTFs by assuming that the auditory system behaves like a low-pass filter that smoothes envelope fluctuations faster than about 50–60 Hz. However, an alternative model—the modulation filterbank model—(e.g., Dau, 1996; Dau *et al.*, 1997a, b; Ewert and Dau, 1999, 2000) assumes that the auditory system decomposes the temporal envelope of sounds via a bank of overlapping bandpass filters, with center frequencies ranging from about 1–100 Hz and bandwidth increasing with center frequency. A number of modulation masking and adaptation experiments (e.g., Kay and Matthews, 1972; Green and Kay, 1974; Tansley and Suffield, 1983; Rees and Kay, 1985; Bacon and Grantham, 1989; Houtgast, 1989; Yost *et al.*, 1989; Dau *et al.*, 1997a, b) support this view. At present, there is controversy as to which model better describes these modulation perception data.

In a recent study (Lorenzi *et al.*, 2001), detection thresholds were measured for a sinusoidal modulation applied to the *modulation depth* of a sinusoidally amplitude-modulated white noise carrier as a function of the frequency of the modulation applied to the modulation depth (referred to as f'_m). The SAM noise acted therefore as a ''carrier'' stimulus of frequency f_m, and sinusoidal modulation of the SAM-noise modulation depth generated two additional components in the modulation spectrum at frequencies $f_m - f'_m$ and $f_m + f'_m$. The tracking variable was the modulation depth of the sinusoidal variation applied to the carrier modulation depth. The resulting ''second-order'' TMTFs measured for four listeners for carrier modulation frequencies f_m of 16, 64, and 256 Hz display a low-pass segment followed by a plateau. These results indicate that, for broadband noise carriers,

auditory sensitivity to regular fluctuations in the strength of amplitude modulation is best for fluctuation rates f'_m below about 2–4 Hz. Parallel measurements of (first-order) TMTFs and masked modulation detection thresholds revealed that this capacity is possibly related to the detection of a beat in the sound's temporal envelope at frequency f'_m. The measured second-order modulation detection thresholds were qualitatively consistent with theoretical predictions of a linear envelope detector model. Unlike listeners' performance, simulations of a linear modulation filterbank model using a Q value of 2 (as in Dau, 1996; Dau *et al.*, 1997a, b) predicted that second-order modulation detection thresholds should decrease at high second-order modulation frequencies f'_m because the modulation sidebands are spectrally resolved at high values of f'_m. This study therefore suggests that, if such modulation filters do exist, their selectivity is extremely poor. The Q value of such modulation filters would have to be less than 2 (in agreement with Ewert and Dau, 1999, 2000), and detection of second-order modulation may then be based on the detection of spectral energy at the envelope beat frequency in the ''internal'' modulation spectrum of the stimuli.

Applying a compressive nonlinearity before envelope extraction yields a low-frequency component in the modulation spectrum of stimuli with second-order SAM. A low-frequency component of this nature may be heard as a beat. This idea is also supported by previous electrophysiological and psychoacoustical studies using two-component modulators (Shofner *et al.*, 1996; Sheft and Yost, 1997; Moore *et al.*, 1999). The compressive nonlinearity may correspond to the fast-acting compression performed by active mechanisms within the cochlea (e.g., Ruggero *et al.*, 1997; Moore and Oxenham, 1998; Nobili *et al.*, 1998) or to the transduction from basilar-membrane vibration to neural activity (e.g., Yates, 1990; Regan and Regan, 1993). Figure 1 illustrates the effect of applying compression to a stimulus with first- and second-order amplitude modulation. The stimulus was a 5-kHz pure tone sinusoidally amplitude modulated at a modulation frequency f_m of 16 Hz. The left panels show the modulation spectrum of this SAM tone when modulation depth m is constant and equal to 50% (i.e., first-order AM). The right panels show the modulation spectrum of this SAM tone when sinusoidal modulation is applied to its modulation depth m at a frequency f'_m of 6 Hz, with a modulation depth m' of 50% (i.e., second-order AM). All stimuli were equated in energy. The bottom and top panels show the modulation spectra of stimuli with and without compression, respectively. In the bottom panels, cochlear compression was simulated by raising the stimuli to the power 0.3 before envelope extraction. In the absence of compression (top panels), the modulation spectrum of the tone with first-order SAM shows a single peak at f_m, and the modulation spectrum of the tone with second-order SAM shows the expected sidebands at $f_m - f'_m$ and $f_m + f'_m$. The bottom panels reveal that compression generates a secondary peak at $2\,f_m$ in the modulation spectrum of the tone with first-order SAM, and a strong component at the envelope beat frequency f'_m (shown by an arrow) in the modulation spectrum of the tone with second-order SAM. Additional distortion components (at $2\,f_m$, for instance) also appear in the high modulation-frequency region of the modulation spectrum. However, the strong component at the envelope beat frequency f'_m might play a critical role in the detection of second-order SAM on the basis that listeners usually report identifying the second-order SAM by detecting a *slow* rhythmic pattern, that is a cyclic increase and decrease in the modulation depth at a rate close to f'_m.

The purpose of the present experiments is to assess to what extent detection of spectral energy at the beat frequency f'_m may account for the specific characteristics of second-order TMTFs. To address this issue, second-order temporal modulation transfer functions (TMTFs) were measured for three listeners using a 5-kHz pure-tone carrier and a narrow-band noise carrier with center frequency and bandwidth of 5 kHz and 2 Hz, respectively. As shown previously by Dau *et al.* (1997a), the inherent statistical fluctuations of the narrow-band 2-Hz-wide noise carrier should mask any component in the low-frequency region ($\leqslant$2 Hz) of the stimuli's modulation spectrum, and this should theoretically prevent listeners from detecting spectral energy at the beat frequency f'_m (or at least, degrade listeners' capacity to make use of such low temporal envelope cues). The modulation filterbank model (as implemented by Dau, 1996; Dau *et al.*, 1997a, b) should therefore predict a loss of second-order modulation sensitivity restricted to the lowest values of f'_m, giving second-order TMTFs a bandpass- or high-pass shape instead of a low-pass one. In contrast to the predictions of the modulation filterbank model, a linear envelope detector model (as implemented by Viemeister, 1979; Strickland and Viemeister, 1996; or Lorenzi *et al.*, 1999) should predict an overall decrease in second-order modulation sensitivity independent of f'_m. First-order TMTFs using the same pure-tone and narrow-band noise carriers were also measured in the three listeners to evaluate the amount and spread of modulation masking produced by the narrow-band noise carrier.

II. METHOD

A. Listeners

Three listeners with normal hearing, CL, MS, and RB (ages ranged from 21 to 32 years), participated in the experiments.

B. Stimuli and procedures

All psychophysical experiments were controlled by a personal computer. All stimuli were generated using a 16-bit D/A converter at a sampling frequency of 44.1 kHz, and were delivered binaurally via Sennheiser HD 565 earphones at a sound-pressure level (SPL) of 80 dB. The narrow-band noises used in all experiments were physically different within and between trials. When pure-tone or narrow-band noise carriers were modulated, modulation was applied over the full duration of the carrier (i.e., a 2-s stimulus duration including 50-ms rise/fall times shaped using a raised-cosine function). In each experiment, the interstimulus interval was 1 s.

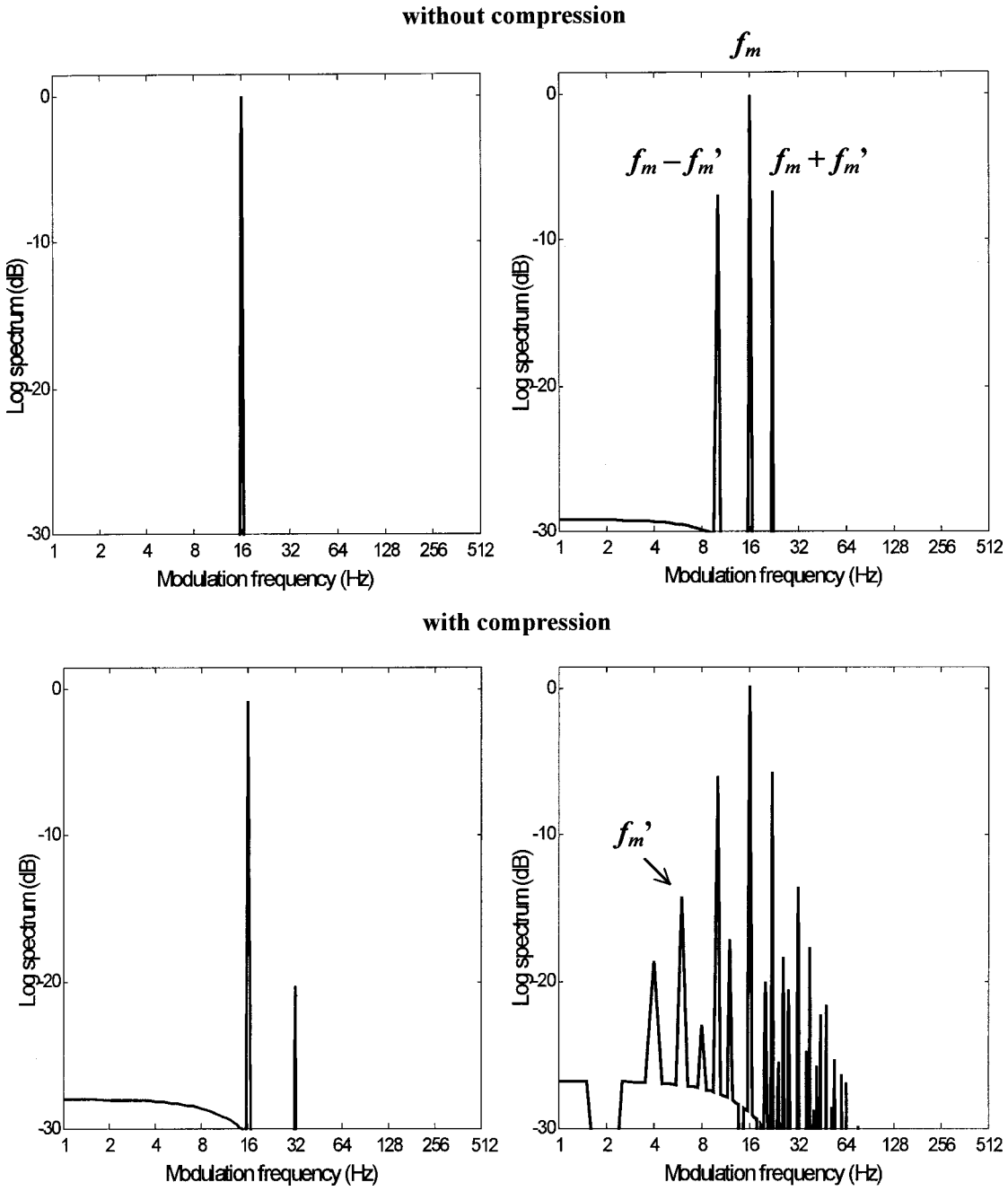


FIG. 1. Top panels: Modulation spectra of a 5-kHz pure tone with first-order (left panel) and second-order (right panel) SAM. No compression is applied to the stimulus before envelope extraction. Bottom panels: Modulation spectra of a 5-kHz pure tone with first-order (left panel) and second-order (right panel) SAM. Compression is applied to the stimulus before envelope extraction. The arrow shows the distortion component at the envelope beat frequency. Stimuli parameters: $f_m = 16$ Hz, $f_m' = 6$ Hz, $m = m' = 0.5$. All stimuli were equated in energy.

1. First-order TMTFs

Listeners had to detect the presence of a SAM applied to a pure tone or to a 2-Hz-wide, narrow-band noise carrier. On each trial, a standard and a target stimulus were successively presented in random order to the listener. The standard, $S(t)$, consisted of either a 5-kHz pure-tone carrier or a narrow-band Gaussian noise centered at 5 kHz, $n(t)$. The noise stimuli were digitally filtered by setting the magnitude of the Fourier coefficients to zero outside the 2-Hz passband. Figure 2 shows the average long-term modulation spectrum of all the narrow-band Gaussian noises used. The average long-term modulation spectrum has a low-pass shape, dropping to -16.7 dB at 2 Hz and to -60 dB at 10 Hz. The target, $T(t)$, consisted of a carrier $n(t)$ similar to that used in the standard (i.e., pure-tone or narrow-band noise) which was sinusoidally amplitude modulated at a given modulation frequency. The expression describing the target was

$$T(t) = c[1 + m \sin(2\pi f_m t + \phi_m)]n(t), \tag{1}$$

where m is the modulation depth ($0 \leqslant m \leqslant 1$), f_m is the modulation frequency (f_m was 1, 2, 4, 8, 16, 23, 32, 64, 128, and 256 Hz), and ϕ_m is the starting phase of the modulation, randomized on each interval. The term c is a multiplicative compensation term (Viemeister, 1979) set such that the over-

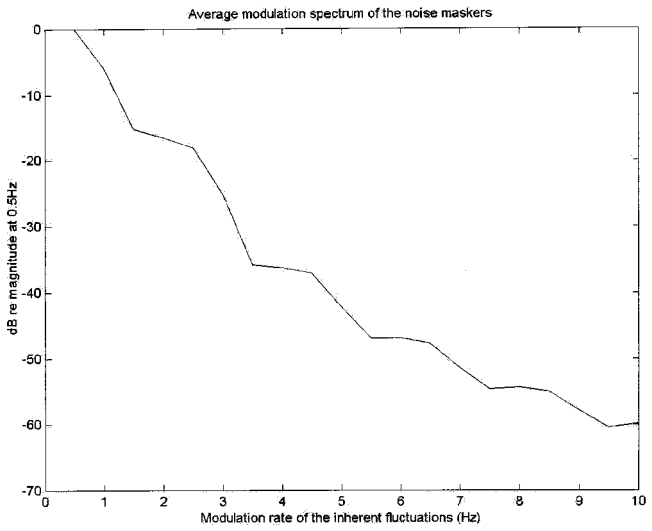


FIG. 2. Long-term modulation spectrum of all the narrow-band Gaussian noises used as carrier signals in the present study. This figure shows that the inherent amplitude fluctuations in the noise occur at low rates.

all power was the same in all intervals. The expression for c is given as follows:

$$c=[1+m^2/2]^{-0.5}. \tag{2}$$

SAM-detection thresholds were obtained using an adaptive two-interval, two-alternative forced-choice (2I, 2AFC) procedure with a two-down, one-up stepping rule that estimates the modulation depth, m, necessary for 70.7%-correct detection. Listeners received visual feedback concerning the correct interval after each trial. The step size of m variation corresponded initially to a factor of 1.585 [4 dB in decibels $(20\log m)$]; it was reduced to 1.26 (2 dB) after the first two reversals. The mean of the last 10 reversals in a block of 16 reversals was taken as the threshold estimate for that block (in %). For each listener and for each modulation frequency, thresholds presented here are based upon three estimates. The worst threshold that can be measured corresponds to a modulation depth of 1 (100% modulated noise). The closer to zero the value of m, the more sensitive the listener is to modulation.

2. Second-order TMTFs

Listeners had to detect the presence of a sinusoidal modulation applied to the modulation depth of a SAM pure-tone carrier or to a SAM 2-Hz-wide, narrow-band noise carrier. On each trial, a standard and a target stimulus were successively presented in random order to the listener. The standard, $S(t)$, consisted of either a 5-kHz pure-tone carrier or a narrow-band Gaussian noise centered at 5 kHz, $n(t)$, sinusoidally amplitude modulated at a given modulation frequency f_m, with a fixed modulation depth m of 50% ($m=0.5$). The expression describing the standard was

$$S(t)=[1+m\sin(2\pi f_m t+\phi_m)]n(t), \tag{3}$$

where ϕ_m represents the starting phase of the modulation, randomized on each interval. The target, $T(t)$, consisted of a carrier similar to that used in the standard (i.e., pure-tone or narrow-band noise). It was amplitude modulated at a given modulation frequency f_m, and its modulation depth was sinusoidally amplitude modulated at a given modulation frequency f'_m. The expression describing the target was

$$T(t)=[1+[m+m'\sin(2\pi f'_m t+\phi'_m)]\times\sin(2\pi f_m t+\phi'_m)]n(t), \tag{4}$$

where m' is the modulation depth of modulation-depth variation ($0\leqslant m'\leqslant 0.5$), m is the "carrier" modulation depth ($m=0.5$), f'_m is the frequency of modulation depth variation, and ϕ'_m represents the starting phase of modulations, randomized on each interval. The overall power was the same in all intervals. f_m (the carrier modulation frequency) was 16, 64, or 256 Hz. f'_m (the "second-order" modulation frequency) was 1, 2, 3, 4, 5, 6, 7, 8, 9, or 11 Hz when f_m was 16 Hz; 1, 2, 3, 4, 6, 8, 11, 16, 23, or 45 Hz when f_m was 64 Hz; and 1, 2, 3, 6, 16, 23, 32, 45, 90, or 181 Hz when f_m was 256 Hz.

Second-order SAM-detection thresholds were obtained using an *identical* psychophysical procedure to that used to gain first-order SAM-detection thresholds (for details, see also Lorenzi *et al.*, in press).

III. RESULTS

A. First- and second-order TMTFs measured with a 5-kHz pure-tone carrier

Individual and mean data for the three listeners are shown in Fig. 3. For comparison, the mean data are replotted in Fig. 4 (left panel) along with the mean data obtained by Lorenzi *et al.* (2001) with four listeners and white noise carries (right panel). In each panel, the second-order TMTFs measured for $f_m=16$ Hz (solid lines with unfilled circles), $f_m=64$ Hz (solid lines with gray squares), and $f_m=256$ Hz (solid lines with black triangles) are plotted along with the first-order TMTF (dotted lines with stars). In the case of second-order TMTFs, the ordinate indicates second-order modulation depth at threshold m', and the abscissa represents f'_m; in the case of first-order TMTFs, the ordinate indicates first-order modulation depth at threshold m, and the abscissa represents f_m.

In agreement with previous studies (e.g., Zwicker, 1952; Viemeister, 1979; Fassel, 1994; Dau, 1996; Kohlrausch, Fassel, and Dau, 2000), first-order TMTFs measured with a pure-tone carrier display a low-pass characteristic. The cutoff frequency is higher than that obtained from first-order TMTFs measured with a broadband noise carrier. On average, SAM-detection thresholds increase slightly but continuously up to $f_m=128$ Hz, and sensitivity is reduced by a factor of 3 (or by about 9 dB in $20\log m$) at $f_m=256$ Hz. For each carrier modulation frequency, second-order TMTFs measured with the pure-tone carrier also display a low-pass shape: sensitivity to second-order modulation decreases slightly but continuously with increasing f'_m. Figure 3 also shows that sensitivity to second-order modulation generally increases with f_m. Sensitivity to second-order modulation approaches sensitivity to first-order modulation when $f_m=256$ Hz. In this case, the second-order TMTF averaged across listeners clearly mirrors the average first-order TMTF.

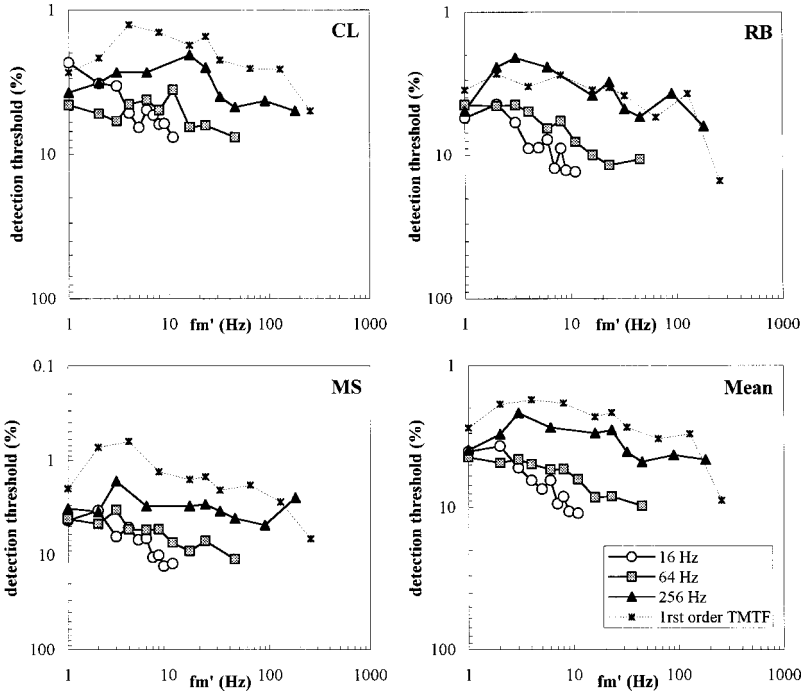


FIG. 3. Individual and mean data for the three listeners. In each panel, the second-order TMTFs measured for $f_m = 16$ Hz (solid lines with unfilled circles), $f_m = 64$ Hz (solid lines with gray squares), and $f_m = 256$ Hz (solid lines with black triangles) are plotted along with the first-order TMTF (dotted lines with stars). In the case of second-order TMTFs, the ordinate indicates second-order modulation depth at threshold m', and the abscissa represents f_m'; in the case of first-order TMTFs, the ordinate indicates first-order modulation depth at threshold m, and the abscissa represents f_m. In each condition, the carrier is a 5-kHz pure tone.

This is especially true for listener RB, whose second-order TMTF measured for $f_m = 256$ Hz overlaps her first-order TMTF.

This contrasts with the data obtained with a white noise carrier (Fig. 4, right panel). In the latter case, sensitivity to second-order modulation degrades above 2 Hz when $f_m = 16$ Hz; 4 Hz when $f_m = 64$ Hz; and 3 Hz (or less) when $f_m = 256$ Hz. Moreover, overall sensitivity to second-order modulation degrades considerably for $f_m = 256$ Hz, compared with sensitivity measured for $f_m = 16$ and 64 Hz.

B. First-order TMTFs measured with a narrow-band noise carrier

Figure 5 shows the individual and mean first-order TMTFs measured with the 5-kHz pure-tone (unfilled circles) and the 2-Hz-wide, narrow-band noise (filled circles) carriers. The present data accurately replicate the detection data of Dau *et al.* (1997a) obtained with a 3-Hz narrow-band noise carrier. For each modulation frequency, the amount of modulation masking produced by the inherent statistical fluctuations of the narrow-band noise carrier was taken as the difference between the unmasked and masked detection thresholds, expressed on a decibel scale ($20 \log m$) (unmasked and masked detection thresholds corresponding to thresholds measured with the 5-kHz pure-tone and the 2-Hz-wide, narrow-band noise carriers, respectively). Maximum modulation masking of 25 dB occurs at $f_m = 1$ Hz; with increasing modulation frequency, the masked detection thresholds decrease and converge with those obtained using a sinusoidal carrier at about $f_m = 32$ Hz. This is consistent with the idea that inherent random fluctuations of the narrow-band 2-Hz-wide carrier mask the additional periodic target modulation.

C. Second-order TMTFs measured with a narrow-band noise carrier

Figures 6, 7, and 8 show the individual and mean second-order TMTFs measured with the 5-kHz pure-tone (unfilled circles) and the 2-Hz-wide, narrow-band noise (filled circles) carriers when the carrier modulation frequency, f_m, is 16, 64, and 256 Hz, respectively. As for first-order TMTFs, the inherent random fluctuations of the narrow-band noise carrier mask the additional periodic (first- and second-order) modulations. Second-order TMTFs measured with the narrow-band noise carrier are nearly flat for $f_m = 16$ and 64 Hz, and show a high-pass segment for f_m

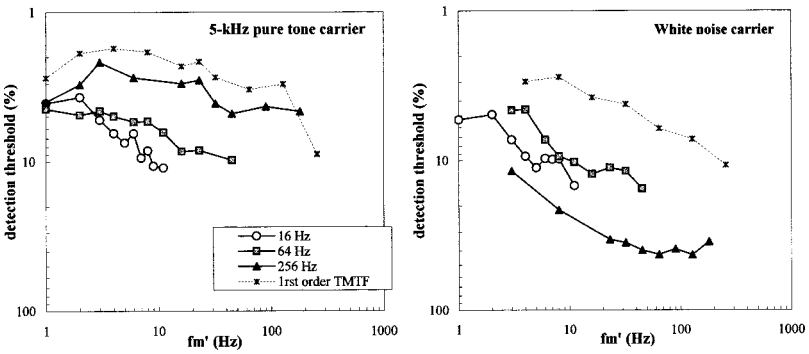


FIG. 4. Left panel: Mean second-order TMTFs for the three listeners; the carrier was a 5-kHz pure tone. Right panel: Mean second-order TMTFs obtained by Lorenzi *et al.* (2001) with four listeners; the carrier was a white noise. See Fig. 3 for legend details.

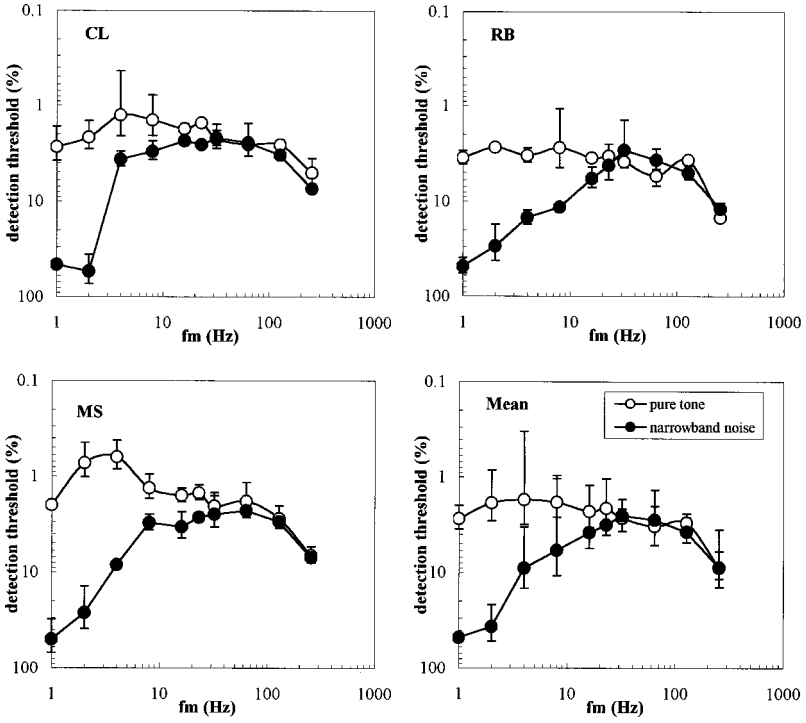


FIG. 5. Individual first-order TMTFs measured for the three listeners using (i) a 5-kHz pure-tone carrier (unfilled circles), and (ii) a 2-Hz-wide narrow-band noise carrier centered at 5 kHz (filled circles). In each panel showing individual data, error bars show one standard deviation about the mean of three threshold estimates. In the panel showing the mean data for the three listeners; error bars show one standard deviation about the mean.

$=256$ Hz. Overall, modulation masking decreases with increasing second-order modulation frequency f'_m.

D. Control experiments

A 5-kHz pure tone was used as a carrier in an attempt to prevent listeners from using spectral cues when performing the second-order modulation detection task, since (audio) frequency selectivity is extremely poor at this carrier frequency. However, it still remains possible that high-frequency cues might have been detected by listeners at the highest values of f_m and f'_m used in the present experiments. Moreover, Wiegrebe and Patterson (1999) recently showed that SAM applied to band limited noise or tone carriers generates audible distortion products in the region of the modulation frequency. It follows that low-frequency cues might also have influenced listeners' performance in the present experiments. To address these issues, second-order TMTFs at $f_m=256$ Hz were measured in two listeners (i) using a 9-kHz pure-tone carrier; (ii) using a 5-kHz pure-tone carrier and adding a low-pass masking noise [Gaussian noise low-pass

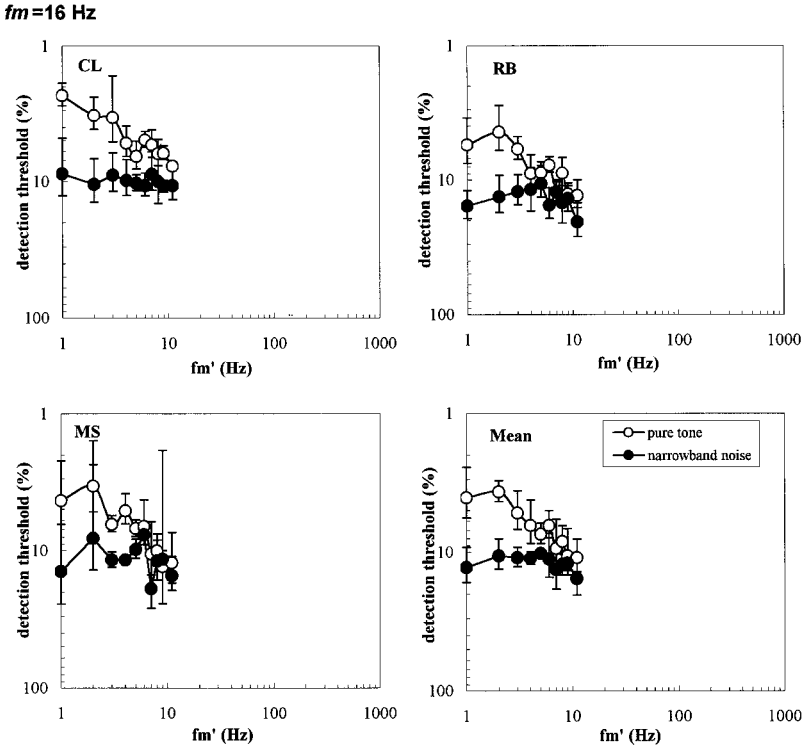


FIG. 6. Individual and mean data for the three listeners. Each panel shows the second-order TMTFs measured for $f_m=16$ Hz using (i) a 5-kHz pure-tone carrier (unfilled circles) and (ii) a 2-Hz-wide narrow-band noise carrier (filled circles). See Fig. 3 for legend details.

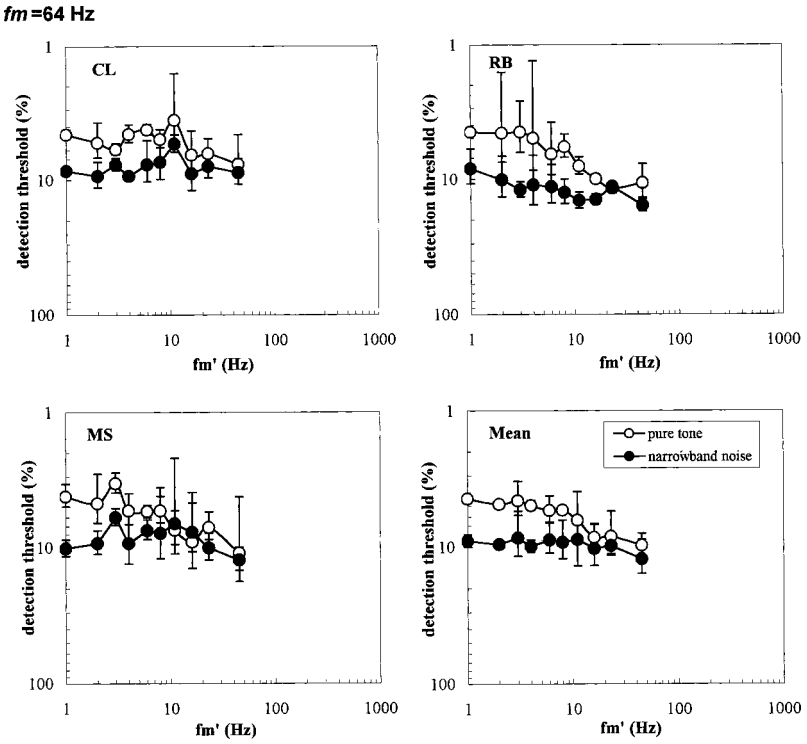


FIG. 7. Individual and mean data for the three listeners. $f_m = 64$ Hz. See Fig. 6 for legend details.

filtered at 500 Hz (12-dB/oct rolloff) presented at 60 dB SPL]. The results are presented in Fig. 9. For comparison, the individual data are plotted along with the individual second-order TMTFs measured previously with a 5-kHz pure-tone carrier (and without any low-pass masking noise). Second-order TMTFs were similar in the three experimental conditions, indicating that the present second-order modulation detection thresholds are not contaminated by the detection of low or high spectral cues.

IV. GENERAL DISCUSSION

The results obtained with a 5-kHz pure-tone carrier show that the overall sensitivity and cutoff frequency estimated from second-order TMTFs increase when the carrier modulation frequency f_m increases from 16 to 256 Hz. At the highest carrier modulation frequency tested (256 Hz), second-order modulation detection thresholds are very close to first-order modulation detection thresholds (i.e., the ''clas-

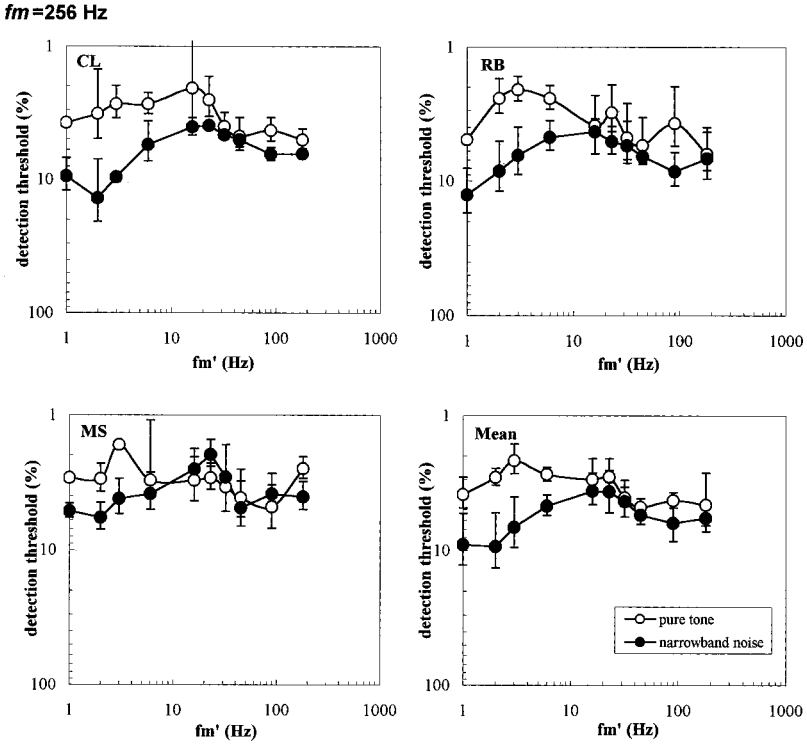


FIG. 8. Individual and mean data for the three listeners. $f_m = 256$ Hz. See Fig. 6 for legend details.

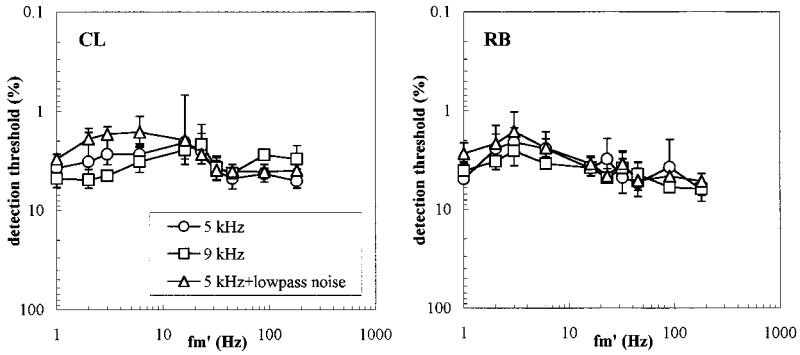


FIG. 9. Individual second-order TMTFs for two listeners, measured for (i) a 5-kHz pure-tone carrier (unfilled circles); (ii) a 9-kHz pure-tone carrier (unfilled squares), and (iii) a 5-kHz pure-tone carrier with a low-pass masking noise (unfilled triangles). Error bars show one standard deviation about the mean of three thresholds estimates.

sic'' TMTF). This contrasts with the substantial drop in second-order modulation sensitivity observed at $f_m = 256$ Hz when using a white noise carrier (Lorenzi *et al.*, in press). In agreement with the results of the study by Dau *et al.* (1997a) on (first-order) SAM detection, the present results suggest that, when using a white noise carrier (i) the intrinsic statistical fluctuations of this noise carrier mask the carrier and sidebands modulations, and (ii) the masking effect is maximum at the highest modulation rate tested, i.e., 256 Hz.

For all listeners, second-order modulation detection thresholds converge at about 4% when $f'_m = 1$ Hz. It may therefore be the case that, at the lowest second-order modulation rate tested (here, $f'_m = 1$ Hz), second-order modulation detection thresholds merely reflect modulation-depth discrimination ability. Overall, these second-order modulation detection thresholds measured at $f'_m = 1$ Hz are consistent with the modulation-depth discrimination thresholds reported in the literature under similar conditions of standard modulation depth and modulation frequency (Fleischer, 1980; Ozimek and Sek, 1988; Wakefield and Viemeister, 1990; Moore and Jorasz, 1992; Lee and Bacon, 1997).

In agreement with our initial predictions of the modulation filterbank model, the results obtained with a 2-Hz-wide, narrow-band noise carrier show that the masking effect produced by the intrinsic statistical fluctuations of the narrow-band noise is mostly restricted to the lowest second-order modulation frequencies ($\leqslant$16 Hz). Moreover, the intrinsic statistical fluctuations of the narrow-band noise carrier do not eliminate completely the ability to detect second-order modulation (the worst thresholds are never greater than about 15%). Taken together, these results suggest that detection of spectral energy at the envelope beat frequency f'_m contributes—at least in part—to second-order modulation detection. The increase in the cutoff frequency of second-order TMTFs with f_m and the similarity of first-order TMTFs and second-order TMTFs measured at $f_m = 256$ Hz suggest that this contribution increases with the spectral separation between f_m (the carrier rate) and f'_m (the beat rate). This may be related to the fact that, at low carrier modulation frequencies (i.e., 16 or 64 Hz), where f_m and f'_m are very close, the lower sideband and carrier modulation components probably mask the component at the envelope beat frequency f'_m. In agreement with previous modulation masking data (e.g., Bacon and Grantham, 1989; Houtgast, 1989; Strickland and Viemeister, 1996; Lorenzi *et al.*, 1997), such a masking effect would disappear when the lower sideband and carrier components are far enough from the envelope beat region.

In summary, the present data show that, under certain conditions, detection of spectral energy at the envelope beat frequency contributes to the detection of second-order modulation. This emphasizes the idea that nonlinear mechanisms in the auditory pathway produce audible effects *both* in the audio-frequency and amplitude-modulation domains (e.g., Shofner *et al.*, 1996; Sheft and Yost, 1997; Moore *et al.*, 1999). However, another (unknown) mechanism seems required to account for our ability to detect second-order modulation, since substantial masking of the beat component does not abolish this ability.

ACKNOWLEDGMENTS

This study was done while C. Lorenzi was a visiting Professor at the Dept. of Physiological Sciences in Newcastle-Upon-Tyne (UK). This work was supported by the Wellcome Trust and the BBSRC. The first author was also supported by the Cognitique program (MENRT).

Bacon, S. P., and Grantham, D. W. (**1989**). ''Modulation masking patterns: Effects of modulation frequency, depth and phase,'' J. Acoust. Soc. Am. **85**, 2575–2580.

Dau, T. (**1996**). ''Modeling auditory processing of amplitude modulation,'' Ph.D. dissertation. Oldenburg: BIS.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). ''Modeling auditory processing of amplitude modulation. I. Modulation detection and masking with narrow-band carriers,'' J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). ''Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration in modulation detection,'' J. Acoust. Soc. Am. **102**, 2906–2919.

Ewert, S. D., and Dau, T. (**1999**). ''Psychophysical tuning in auditory AM-processing,'' in *Psychophysics, Physiology and Models of Hearing* (World Scientific, Singapore), pp. 73–76.

Ewert, S. D., and Dau, T. (**2000**). ''Characterizing frequency selectivity for envelope fluctuations,'' J. Acoust. Soc. Am. **108**, 1181–1196.

Fassel, R. (**1994**). ''Experimente und Simulationsrechnungen zur Wahrnehmung von Amplitudenmodulationen im menschlichen Gehör,'' Ph.D. dissertation. University of Göttingen.

Forrest, T. G., and Green, D. M. (**1987**). ''Detection of partially filled gaps in noise and the temporal modulation transfer function,'' J. Acoust. Soc. Am. **82**, 1933–1943.

Green, G. G. R., and Kay, R. H. (**1974**). ''Channels in the human auditory system concerned with the waveform of the modulation present in amplitude-and frequency-modulated tones,'' J. Physiol. (London) **241**, 29–30P.

Houtgast, T. (**1989**). ''Frequency selectivity in amplitude-modulation detection,'' J. Acoust. Soc. Am. **85**, 1676–1680.

Kay, R. H., and Matthews, D. R. (**1972**). ''On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones,'' J. Physiol. (London) **225**, 657–677.

Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). ''The influence of carrier level and frequency on modulation and heat-detection thresholds for sinusoidal carriers,'' J. Acoust. Soc. Am. **108**, 723–734.

Lee, J., and Bacon, S. P. (**1997**). ''Amplitude modulation depth discrimination of a sinusoidal carrier: Effect of stimulus duration,'' J. Acoust. Soc. Am. **101**, 3688–3693.

Lorenzi, C., Berthommier, F., and Demany, L. (**1999**). ''Discrimination of amplitude-modulation phase spectrum,'' J. Acoust. Soc. Am. **105**, 2987–2990.

Lorenzi, C., Micheyl, C., Berthommier, F., and Portalier, S. (**1997**). ''Modulation masking in listeners with sensorineural hearing loss,'' J. Speech Hear. Res. **40**, 200–207.

Lorenzi, C., Soares, C., and Vonner, T. (**2001**). ''Second-order temporal modulation transfer functions,'' J. Acoust. Soc. Am. **110**, 1030–2038.

Moore, B. C. J., and Jorasz, U. (**1992**). ''Detection of changes in modulation depth of a target sound in the presence of other modulated sounds,'' J. Acoust. Soc. Am. **91**, 1051–1061.

Moore, B. C. J., and Oxenham, A. J. (**1998**). ''Psychoacoustic consequences of compression in the peripheral auditory system,'' Psychol. Rev. **105**, 108–124.

Moore, B. C. J., Sek, A., and Glasberg, B. R. (**1999**). ''Modulation masking produced by beating modulators,'' J. Acoust. Soc. Am. **106**, 908–918.

Nobili, R., Mammano, F., and Ashmore, J. (**1998**). ''How well do we understand the cochlea?'' Trends Neurosci. **21**, 159–167.

Ozimek, E., and Sek, A. (**1988**). ''AM difference limens for noise bands,'' Acustica **66**, 153–160.

Rees, A., and Kay, R. H. (**1985**). ''Delineation of FM rate channels in man by detectability of a three component modulation waveform,'' Hear. Res. **18**, 211–221.

Regan, M. P., and Regan, D. (**1993**). ''Nonlinear terms produced by passing amplitude-modulated sinusoids through a hair cell transducer function,'' Biol. Cybern. **69**, 439–446.

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.

Sheft, S., and Yost, W. A. (**1997**). ''Modulation detection interference with two-component masker modulators,'' J. Acoust. Soc. Am. **102**, 1106–1112.

Shofner, W. P., Sheft, S., and Guzman, S. J. (**1996**). ''Responses of ventral cochlear nucleus units in the chinchilla to amplitude modulation by low-frequency, two-tone complexes,'' J. Acoust. Soc. Am. **99**, 3592–3605.

Strickland, E. A., and Viemeister, N. F. (**1996**). ''Cues for discrimination of envelopes,'' J. Acoust. Soc. Am. **99**, 3638–3646.

Tansley, B. W., and Suffield, J. B. (**1983**). ''Time-course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation,'' J. Acoust. Soc. Am. **74**, 765–775.

von Fleischer, H. (**1980**). ''Subjektive grösse von unterschieden im amplituden-modulationsgrad von sinustönen,'' Acustica **46**, 31–37.

Viemeister, N. F. (**1979**). ''Temporal modulation transfer functions based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.

Wakefield, G. H., and Viemeister, N. F. (**1990**). ''Discrimination of modulation depth of SAM noise,'' J. Acoust. Soc. Am. **88**, 1367–1373.

Wiegrebe, L., and Patterson, R. D. (**1999**). ''Quantifying the distortion products generated by amplitude-modulated noise,'' J. Acoust. Soc. Am. **106**, 2709–2718.

Yates, G. K. (**1990**). ''Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions,'' Hear. Res. **50**, 145–162.

Yost, W. A., Sheft, S., and Opie, J. (**1989**). ''Modulation interference in detection and discrimination of amplitude modulation,'' J. Acoust. Soc. Am. **86**, 2138–2147.

Zwicker, E. (**1952**). ''Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones,'' Acustica **2**, 125–133.

Perceptual fusion and fragmentation of complex tones made inharmonic by applying different degrees of frequency shift and spectral stretch

Brian Roberts[a] and Jeffrey M. Brunstrom[b]
School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, England

(Received 5 February 2001; revised 14 August 2001; accepted 17 August 2001)

Global pitch depends on harmonic relations between components, but the perceptual coherence of a complex tone cannot be explained in the same way. Instead, it has been proposed that the auditory system responds to a common pattern of equal spacing between components, but is only sensitive to deviations from this pattern over a limited range [Roberts and Brunstrom, J. Acoust. Soc. Am. **104**, 2326–2338 (1998)]. This hypothesis predicts that spectral fusion will be largely unaffected either by frequency shifting a harmonic stimulus (because equal spacing is preserved), or by small degrees of spectral stretch (because significant deviations from equal spacing only cumulate over large spectral distances). Complex tones were either shifted by 0%–50% of $F0$ (200 Hz±10%) or stretched by 0%–12% of $F0$ (100 Hz±10%). Subjects heard a complex followed by a pure tone in a continuous loop. One of the components 2–11 was mistuned by ±4%, and subjects adjusted the pure tone to match its pitch. Broadly consistent with our hypothesis, frequency shifts had relatively little effect on hit rates and only large degrees of stretch reduced them substantially. The implications for simultaneous grouping are explored with reference to an autocorrelation model of auditory processing. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1410965]

PACS numbers: 43.66.Ba, 43.66.Fe, 43.66.Hg [SPB]

I. INTRODUCTION

The auditory grouping of concurrent spectral components has generally been explained in terms of harmonic relations (e.g., Bregman, 1990; Darwin and Carlyon, 1995), and sometimes as a byproduct of global-pitch perception (e.g., Hartmann, 1996). Roberts and Brunstrom (1998) reviewed the literature on the perception of complex tones with various kinds of spectral regularity and noted that harmonicity could not provide a clear account of the reported findings. Three points were of particular note. First, complex tones made inharmonic by logarithmic stretching of their spectra often sounded surprisingly fused (Slaymaker, 1970; Mathews and Pierce, 1980; Cohen, 1980). Second, there is a mismatch between the amount of mistuning required for a harmonic to be heard out as a separate tone (±1.3%–2.1%; Moore *et al.*, 1986) and for it to begin to be excluded from the computation of global pitch (±3%; Moore *et al.*, 1985). Third, a single even harmonic added to an otherwise odd-harmonic complex is typically more salient than its odd neighbors (Roberts and Bregman, 1991; Roberts and Bailey, 1996a, b; Roberts, 1998). This has been interpreted as reflecting a disruption in the regular pattern of spectral spacing defined by the odd harmonics. Together, these findings suggest that auditory grouping processes can utilize spectral regularity in a more general sense than harmonic relations alone (Roberts and Bailey, 1996b).

Roberts and Brunstrom (1998) explored this suggestion further by measuring the ability of listeners to match the pitch of a component that deviated from its original frequency in a regular but inharmonic stimulus. By analogy with the harmonic case, we refer to this deviation as mistuning. The stimuli used were either frequency-shifted or spectrally stretched complex tones. Frequency shifting involves modifying a harmonic series by the addition of a fixed increment to the frequency of each component. These frequency shifts are defined in terms of a percentage of the original fundamental ($F0$) frequency. This kind of spectral manipulation was originally employed in studies of global-pitch perception (de Boer, 1976; Patterson, 1973; Schouten *et al.*, 1962). Spectral stretch involves modifying a harmonic series by the addition of a cumulative increment to component spacing (defined, again, as a percentage of $F0$) with increasing component number (e.g., Fine and Moore, 1993; Roberts and Bailey, 1996b). It was found that listeners performed almost as well at matching a mistuned component in a 15%-shifted complex or a 3%-stretched complex as in a harmonic complex. Roberts and Brunstrom (1998) reasoned that this could not happen unless the shifted and stretched stimuli were well fused perceptually, as otherwise nontarget components would have become salient and hence been matched in error. The characteristic mismatch between expected and measured pitch for a mistuned component, first observed for harmonic stimuli by Hartmann *et al.* (1990), was also found for these shifted and stretched stimuli. Roberts and Brunstrom (1998) used Scheffers' (1983) harmonic-sieve model to demonstrate that a single harmonic template could not account for their findings. Instead, they put forward the idea of an auditory-grouping template that is sensitive to the regular spectral structure of shifted and stretched stimuli.

Roberts and Brunstrom (1998, p. 2336) proposed that this auditory-grouping template was sensitive to deviations

[a]Electronic mail: b.roberts@bham.ac.uk
[b]Current address: Department of Human Sciences, Loughborough University, Loughborough, Leicestershire LE11 3TU, England.

from equal spacing between consecutive components, but only within a limited range. Frequency shifting a harmonic stimulus by 15% of $F0$ does not alter the intercomponent spacing, and so should not alter the ability of the template to detect a single mistuned component. Although a spectral stretch of 3% does result in substantial changes in intercomponent spacing, deviations from equal spacing only cumulate over large spectral distances, and so should not be detectable by a limited-range process. In contrast, a local deviation from equal spacing produced by mistuning a single component should be detectable. The idea of a limited range of operation for a grouping template is supported by the data of Lin and Hartmann (1998) and of Brunstrom and Roberts (1998), who investigated the auditory organization of harmonic stimuli. Roberts and Brunstrom (1998) suggested that the overall strength of perceptual fusion for a complex tone emerges from the ''chaining'' together across the spectrum of a series of such limited-range computations.

The proposal of Roberts and Brunstrom (1998) predicts that any degree of frequency shift will be tolerated by the auditory system, because this manipulation preserves a regular pattern of equal component spacing across the spectrum. The proposal also predicts that spectral stretch will be tolerated only if it does not produce significant deviations from equal component spacing between neighboring components. Therefore, in experiments 1 and 2, we explored the effects of different degrees of frequency shift and spectral stretch, respectively, on the ability of listeners to match a mistuned target. The mistakes that listeners make when attempting to match a mistuned target can themselves be informative, but, in previous studies, they have never been explored in any detail and usually have not been reported at all. Therefore, the pattern of errors made by the listeners was explored as well as the successful matches. Knowing which nontarget components listeners often match should illuminate the nature of any perceptual fragmentation that develops as the spectral pattern of a complex tone is changed by shift and stretch manipulations.

II. EXPERIMENT 1

A. Method

1. Subjects

Three subjects participated, all of whom reported normal hearing. One subject was the second author; the others were research students from other laboratories in the School of Psychology. These other subjects both had experience of similar analytical listening tasks, having previously taken part in the experiments reported by Roberts and Brunstrom (1998).

2. Task

Each trial consisted of a repeating cycle of a complex tone followed by an adjustable pure tone, with each sound separated from the next by a brief silent interval. One of the components of the complex tone was mistuned from its appropriate spectral-pattern position. The subjects' task was to adjust the frequency of the pure tone until its pitch matched that of the mistuned target. We assume that the ease with which this was possible depended on the perceptual salience of the mistuned component in relation to the other partials in the complex. There was no limit on the number of stimulus cycles that was permitted in a trial. The adjustment was made using a trackball mouse with two different sensitivity settings. The default was a coarse control that allowed rapid adjustment to the chosen frequency region. The fine control was ten times more sensitive, and subjects were encouraged to use this setting before selecting their best match. Subjects chose when to start each trial, and were free to rest at any time between trials.

3. Stimuli and conditions

Within a trial, each stimulus cycle contained a 420-ms complex tone followed by a 310-ms pure tone, both including linear onset and offset ramps of 20 ms each. The durations of the within- and between-cycle silent intervals were 200 and 500 ms, respectively. These were chosen to create a rhythm that clearly identified the within-cycle order of the tones. The complex tone consisted of the first 12 partials of a nominal $F0$, chosen randomly from a rectangular distribution with a width of $\pm 10\%$ around 200 Hz. All of the components were in sine phase and of equal amplitude, each set to 60 dB SPL. The adjustable pure tone was set to the same level. The initial frequency of the adjustable tone was chosen randomly within a fixed range, 100–2800 Hz, quantized in steps of 2 Hz. This range was chosen to exceed that spanned by components 1 and 12 for the lowest and highest possible $F0$ frequencies, respectively, for all conditions. One of 10 components of the complex tone, chosen from the set 2–11, was mistuned by $\pm 4\%$ from its appropriate frequency. Components 1 and 12, which constituted the spectral edges of the complex, were not tested.

The experiment comprised five conditions, which differed in the degree of frequency shift applied to an originally harmonic complex. For conditions 1–5, the frequency shifts used were 0%, 12.5%, 25.0%, 37.5%, and 50.0% of the nominal $F0$, respectively. For example, when the $F0$ of the precursor was 200 Hz, the frequency increment added to each component was 0, 25, 50, 75, and 100 Hz for conditions 1–5, respectively. Note that a frequency shift of 50% transforms a consecutive-harmonic series into an odd-harmonic series with a missing $F0$ component (e.g., 300, 500, 700 Hz, etc.). For ease of comparison across conditions, component frequencies were also represented on a scale comprising units of component number. For the frequency-shifted stimuli used here, one unit on this scale is equivalent to the nominal $F0$ in Hz.

All stimuli were generated using MITSYN software (see Henke, 1990) at a sampling rate of 16 kHz, and played with a precision of 16 bits via a D/A converter (Data Translation DT2823). They were low-pass filtered (corner frequency $=5.2$ kHz, roll-off$=100$ dB/oct) and presented diotically over Sennheiser HD 480-13II earphones. The levels of the stimuli were set using a programmable attenuator, and were calibrated with a sound-level meter (Brüel and Kjaer, Type 2209, linear weighting) connected to the earphones by an

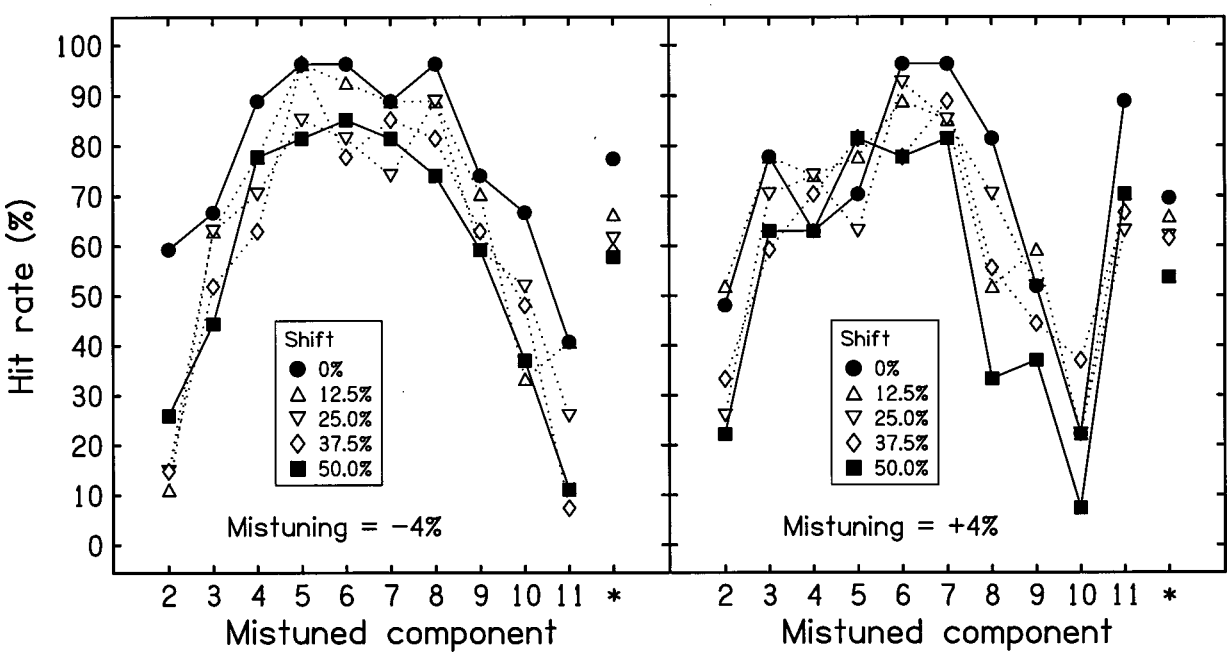


FIG. 1. Mean hit rates for three subjects across degree of frequency shift and component number in experiment 1. The corresponding overall mean hit rates for each condition, collapsed across component number, are indicated by an asterisk on the abscissa. The results for negative and positive mistunings are shown separately in the left-hand and right-hand panels, respectively. Symbols identifying the results for each degree of frequency shift are shown in the panel insets.

artificial ear (Type 4153). The stimuli were presented to the listeners in a sound-attenuating chamber (Industrial Acoustics 1201A).

4. Procedure

Each session consisted of all combinations of component number, direction of mistuning, and condition, giving a total of 100 trials (10 components $\times$ 2 directions $\times$ 5 conditions). The order of these trials was randomized anew for each session. Nine sessions were run per subject, giving a grand total of 900 trials. At the end of each trial, subjects were given feedback on their performance. If the chosen response was within $\pm 1.0 \times$ nominal $F0$ of the correct value, the message ''within range'' was displayed. If the response was above or below this range, the message ''much too high'' or ''much too low'' was displayed, as appropriate. Subjects received a practice session beforehand to familiarize them with the full set of stimuli. This session was the same as a single session in the main experiment.

5. Data analysis

The main analysis was the same as that used by Roberts and Brunstrom (1998). It was developed from a procedure devised by Hartmann *et al.* (1990), who introduced a classification of responses as ''hits'' or ''misses'' that took account of pitch shifts (systematic mismatches between the pitch of a partial and an isolated pure tone of the same frequency). To be classified as a hit, a response was first required to fall within $\pm 1.0 \times$ component-number units from the frequency of the mistuned partial. The second criterion was based on the clustering of the responses that passed the first criterion. A mean and standard deviation were calculated for these data. If the standard deviation exceeded 2.5% of the frequency of the mistuned component, then the data point most distant from the mean was rejected and a new mean and standard deviation were calculated. This procedure was repeated until the clustering criterion was met, or only two data points remained. In cases where these two data points failed the clustering criterion, the data point further from the target frequency was rejected. In cases where only one data point passed the first criterion, that point was accepted as a hit. For each subject, the number of hits per stimulus (0–9) was converted to a percentage. A mean hit rate was then calculated for each stimulus. The distribution across the spectrum of errors made by listeners was also explored (see Secs. II B 2 and III B 2 for details).

B. Results and discussion

1. Hit rates

Figure 1 shows the effects on hit rate of component number and direction of mistuning for each condition. These data were assessed using a three-way repeated-measures analysis of variance (ANOVA). Figure 1 also shows the mean hit rates for each condition when collapsed across component number. Clearly, the listeners were generally well able to hear out mistuned components from both harmonic- and frequency-shifted stimuli.[1] There was a progressive decline in overall performance of about 18% as the degree of shift was increased from 0% to 50% of $F0$ [$F(4,8) = 18.76$, $p < 0.001$]. The hit-rate profiles displayed in Fig. 1 are similar to those observed by Roberts and Brunstrom (1998) for comparable stimuli (harmonic and 15%-shifted). There was a significant main effect of component number [$F(9,18) = 18.20$, $p < 0.001$], which reflects the general tendency for hit rates to be greatest for the middle range of component numbers tested. Although there was no main ef-

fect of direction of mistuning [$F(1,2)=0.14$, $p>0.5$], the interaction of direction $\times$ component was significant [$F(9,18)=4.26$, $p=0.004$]. This reflects the clear tendency for positive-mistuning profiles to peak at components 6 and 7, and to be high for component 11, but for negative-mistuning profiles to be broadly flat over the range 4–9. None of the other interaction terms approached significance.

It is noteworthy that the hit-rate profiles are very similar in overall form across conditions, such that the modest decline in hit rate with increasing degree of frequency shift is the only clear difference between them. This similarity is consistent with the demonstration by Brunstrom and Roberts (2000) that the perceptual fusion of spectral components is not governed by the pitch mechanism (contrary to the argument presented by Hartmann, 1996). The application of an increasing degree of frequency shift to a harmonic complex tone has two effects on its global pitch (Patterson, 1973). First, the global pitch rises in relation to the nominal $F0$ and becomes progressively weaker. Second, another global pitch emerges near the suboctave of the nominal $F0$ and becomes progressively stronger. For our stimuli, the global pitch was most ambiguous for shifts of 25.0% and 37.5%. A shift of 50.0% leads to an odd-harmonic complex with a global pitch an octave below that of the precursor stimulus. Therefore, conditions 1 and 5 shared as common factors both harmonicity and a low ambiguity in global pitch. Despite this, the decline in hit rate was neither reversed nor halted when the degree of shift was increased from 37.5% to 50.0%, as would be expected if either factor was important in determining spectral fusion. Nonetheless, the proposal by Roberts and Brunstrom (1998) of an auditory-grouping template that is sensitive to the equal or near-equal spacing between neighboring partials cannot account fully for the findings, otherwise there would be no decline in performance across conditions.

The basis of the difference in hit-rate profiles for positive and negative mistunings remains unclear, except in the case of component 11. Roberts and Brunstrom (1998) pointed out that the upward mistuning of component 11 brings its frequency close to that of component 12, which forms the upper spectral edge of the complex tone. This component generates a salient edge pitch (Kohlrausch and Houtsma, 1992; Moore and Ohgushi, 1993), and matches to it are generally not distinguishable from matches to the mistuned target. The peak region in the hit-rate profiles was higher than one would predict from the dominant region for an $F0$ of 200 Hz (components 3–5; Ritsma, 1967). This discrepancy was also observed by Roberts and Brunstrom (1998) and can be seen in the data of Hartmann *et al.* (1990) for comparable harmonic stimuli. There is further discussion of this point in Sec. III B 1.

2. Error frequencies

Figure 2 shows the distribution of errors for each condition on occasions when the match failed to meet the first criterion in the hits classification procedure (i.e., it was more than $\pm 1.0 \times$ component-number units from the frequency of the mistuned target). Histogram bins depicting errors falling within ± 0.1 units of a nontarget component are shown filled.

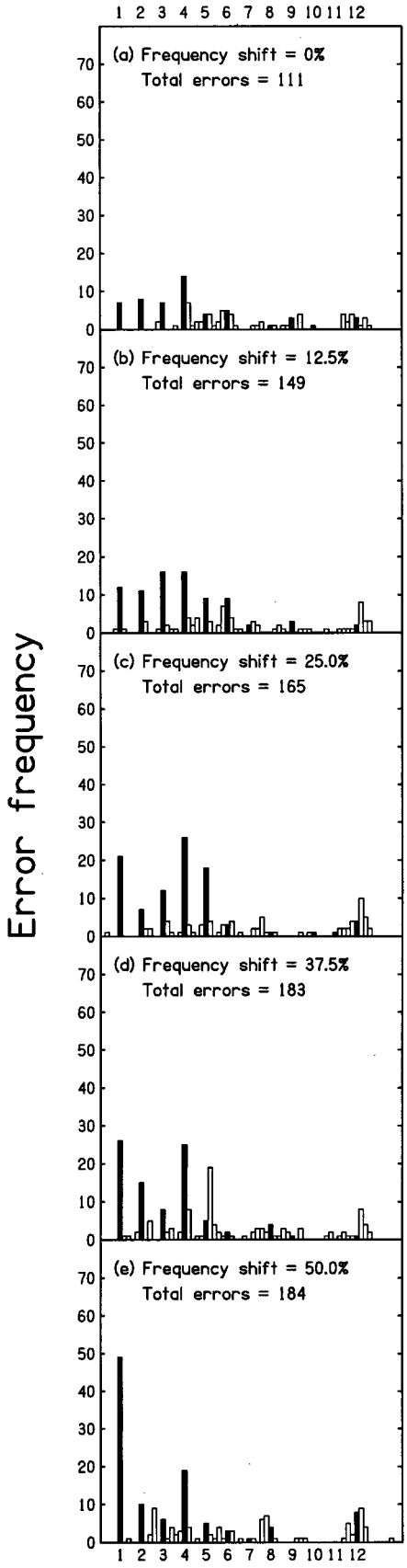


FIG. 2. Cumulative error frequencies for three subjects in experiment 1. The results for each degree of frequency shift are shown separately in panels (a)–(e). Each histogram bin is 0.2 units of component number in width, spanning the range 0.1–14.1. Error frequencies in histogram bins centered on the in-tune frequencies of components 1–12 are shown in black. Each panel also gives the total number of errors when collapsed across all histogram bins.

For all conditions, note that most errors were matches to the resolved components, 1–6. There was also a small cluster of responses around component 12, which corresponds to the upper spectral edge of the stimuli. The effect of increasing the degree of frequency shift was primarily to increase the likelihood of a match to a low-numbered nontarget, especially to component 1. For the maximum shift used, 50.0%, over a quarter of all errors were matches to this component.

Roberts and Bailey (1996b) observed that frequency shifting an odd-harmonic series raised the perceived clarity of its low-numbered components, especially the lowest component. For the greatest degree of frequency shift used in their study (30%), the mean rating for component 1 rose by about one unit on a clarity scale ranging from 1–7. This finding is consistent with an increase in error rate with increasing shift, observed for component 1 in our experiment. Of particular interest is the apparently high salience of component 1 in the 50%-shifted condition. This partial is, in effect, the third harmonic of an odd-harmonic series with a missing $F0$ component. An in-tune harmonic should not attract matches by acting as a salient component, which suggests that harmonicity was not a key factor in determining its integration into the percept of the complex tone.

III. EXPERIMENT 2

Roberts and Brunstrom (1998) suggested that the auditory system is not sensitive to long-range deviations from a pattern of equal spacing between consecutive components. Therefore, moderate degrees of spectral stretch should not lead to the perceptual fragmentation of a complex tone. In contrast, the mistuning of a target component by $\pm 4\%$ causes a significant local deviation from equal spacing that can be utilized to segregate it from the rest of the complex. This leads to the prediction that significant local deviations from equal spacing produced by an appropriate increase in the degree of stretch will cause the complex to decohere. This effect should reduce the odd-one-out status of the mistuned target, making it more difficult to match.

A. Method

Except where stated, the method was the same as used in experiment 1. Experiment 2 comprised five conditions, which differed in the degree of spectral stretch applied to a precursor harmonic complex. For conditions 1–5, the spectral stretches used were 0%, 3%, 6%, 9%, and 12% of the nominal $F0$, respectively. The nominal $F0$ was chosen randomly from a rectangular distribution with a width of $\pm 10\%$ around 100 Hz. For example, when the $F0$ of the original complex was 100 Hz, the cumulative increment added with increasing component number to the frequency spacing of the components was 0, 3, 6, 9, and 12 Hz for conditions 1–5, respectively. The frequencies of the partials comprising the complex tones for a nominal $F0$ of 100 Hz are listed in Table I.

The nominal $F0$ used here was an octave lower than that used in experiment 1. This overcomes the problems caused by the considerable increase in the absolute frequencies of the upper partials that results from using high degrees of spectral stretch. The change in nominal $F0$ ensured that component frequencies always fell below 2.2 kHz, which was the value above which Hartmann *et al.* (1990) observed a progressive decline in the ability of listeners to hear out a mistuned harmonic from an otherwise periodic complex tone. This decline, which probably reflected a reduction in the precision of neural synchrony for partials at higher frequencies (e.g., Anderson, 1973), would otherwise have confounded our attempt to assess the effect of increasing the degree of spectral stretch on listeners' sensitivity to component mistuning. The change in nominal $F0$ also allowed an assessment of the effect of absolute frequency in the harmonic condition, by comparison with its counterpart in experiment 1. The fixed range of adjustment for the pure tone (50–2300 Hz) was chosen to encompass the full range of component frequencies.

Once again, component frequencies were represented on a scale comprising units of component number. For the stretched stimuli used here, the frequency in Hz corresponding to one unit on this scale increased both with component number and degree of stretch. A component-number scale was used to preserve as close an equivalence as possible across conditions for corresponding component numbers, and to facilitate comparison with experiment 1.

B. Results and discussion

1. Hit rates

Figure 3 shows the effects on hit rate of component number and direction of mistuning for each condition. These data were assessed using a three-way repeated-measures ANOVA. Figure 3 also shows the mean hit rates for conditions 1–5 when collapsed across component number. It is clear that listeners were generally able to hear out mistuned components from harmonic stimuli and from stimuli with modest degrees of spectral stretch. However, this was not the case for stimuli with higher degrees of stretch. There was a substantial and progressive decline in overall performance of about 33% as the degree of stretch was increased from 0% to 12% of $F0$ [$F(4,8)=10.51$, $p=0.003$]. There was a significant main effect of component number [$F(9,18)=5.24$, $p=0.001$], which reflects the tendency for hit rates to peak in the region of components 7–9. Except for a decline in magnitude, the hit-rate profiles remain broadly similar in shape across conditions. As in experiment 1, there was no main effect of direction of mistuning [$F(1,2)=0.12$, $p>0.5$], but the interaction of direction $\times$ component was significant [$F(9,18)=7.77$, $p<0.001$]. Once again, the peak in performance for positive mistunings of component 11 was an important contributor to this effect. None of the other interaction terms approached significance.

The results confirm Roberts and Brunstrom's (1998) finding that a small degree of spectral stretch (3%) has little effect on the ability of subjects to match a mistuned component, despite the considerable change in component spacing across the whole spectrum. In contrast, high degrees of stretch cause a marked decline in performance. Indeed, overall performance in the 12%-stretched condition (about 27%) was not greatly above chance.[1] These findings are consistent

TABLE I. For a nominal $F0$ frequency of 100 Hz, columns 2–6 show the in-tune frequencies (in Hz) of the components comprising the complex tones used in experiment 2.

Component No.	Condition 1 (harmonic)	Condition 2 (3%-stretched)	Condition 3 (6%-stretched)	Condition 4 (9%-stretched)	Condition 5 (12%-stretched)
1	100	100	100	100	100
2	200	203	206	209	212
3	300	309	318	327	336
4	400	418	436	454	472
5	500	530	560	590	620
6	600	645	690	735	780
7	700	763	826	889	952
8	800	884	968	1052	1136
9	900	1008	1116	1224	1332
10	1000	1135	1270	1405	1540
11	1100	1265	1430	1595	1760
12	1200	1398	1596	1794	1992

with the hypothesis that significant deviations from equal spacing within local clusters of components are required to decohere a complex tone.

The hit-rate profile for the harmonic condition clearly differs from its counterpart in experiment 1. In particular, the overall hit rate for this condition was about 13% lower than in experiment 1, and the peak region in the profile was shifted upward to components 7–9. This indicates that the tendency for hit-rate profiles to peak in this region, evident even for 12% stretch, is most probably a consequence of the change in nominal $F0$ from 200 to 100 Hz. Consistent with this hypothesis are reports that the dominant region for pitch also moves upward as $F0$ is lowered (Plomp, 1967; Patterson and Wightman, 1976; Moore and Glasberg, 1988). However, as for experiment 1, the peak region in the hit-rate profiles encompasses higher-numbered components than one would predict on the basis of the dominant region.

2. Error frequencies

Figure 4 shows the distribution of errors for each condition on occasions when the match failed to meet the first criterion in the hits classification procedure (i.e., it was more than $\pm 1.0 \times$ component-number units from the frequency of the mistuned target). As for experiment 1, histogram bins depicting errors falling within ± 0.1 units of a nontarget component are shown filled. There is a clear tendency for errors to cluster in the region of component 12 in all conditions. This component represents the upper spectral edge of the complex, which, as noted earlier, is known to evoke a salient pitch. In the harmonic condition, the matches in this region are clearly biased toward the upper side of component 12. Kohlrausch and Houtsma (1991, 1992) made a similar observation, noting that the individual means of their listeners for matches to the 2-kHz spectral edge of a sine-phase harmonic

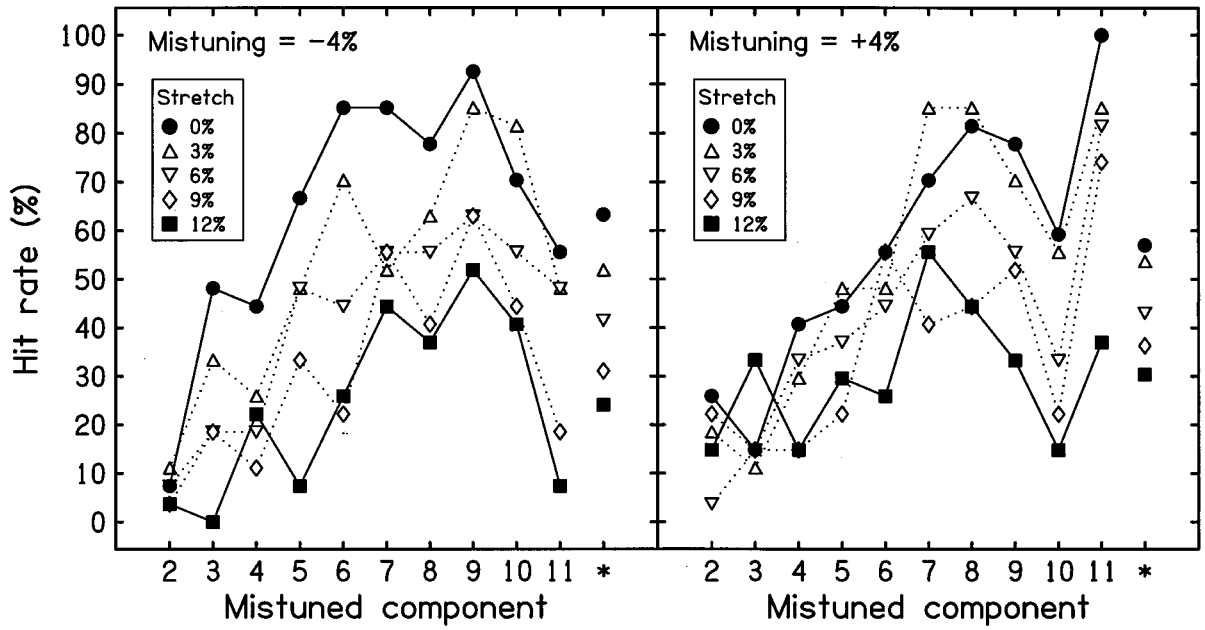


FIG. 3. Mean hit rates for three subjects across degree of spectral stretch and component number in experiment 2. The corresponding overall mean hit rates for each condition, collapsed across component number, are indicated by an asterisk on the abscissa. The results for negative and positive mistunings are shown separately in the left-hand and right-hand panels, respectively. Symbols identifying the results for each degree of spectral stretch are shown in the panel insets.

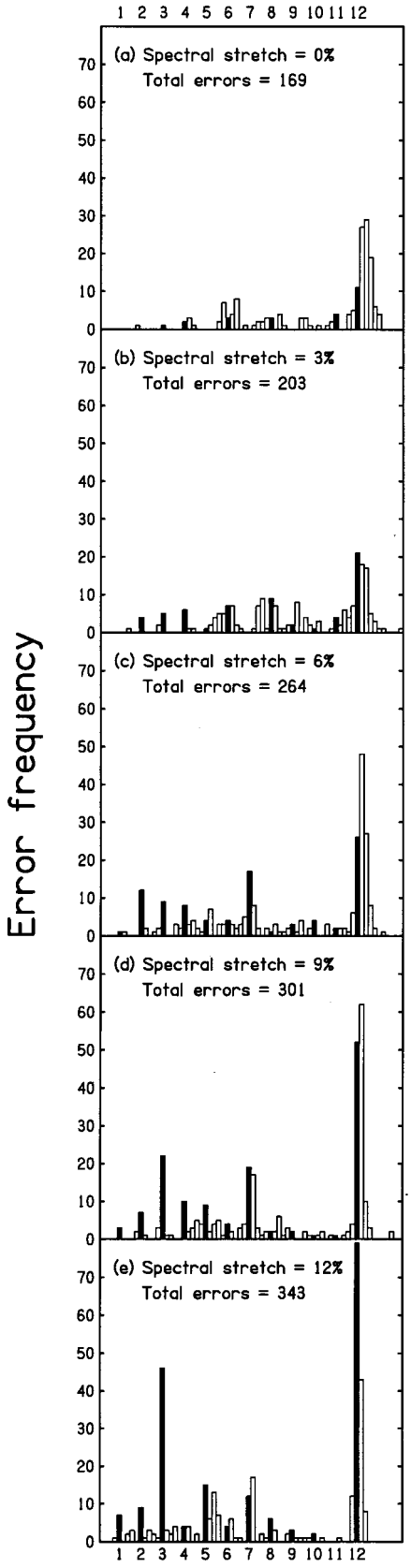


FIG. 4. Cumulative error frequencies for three subjects in experiment 2. The results for each degree of spectral stretch are shown separately in panels (a)–(e). Each histogram bin is 0.2 units of component number in width, spanning the range 0.1–14.1. Error frequencies in histogram bins centered on the in-tune frequencies of components 1–12 are shown in black. Each panel also gives the total number of errors when collapsed across all histogram bins.

complex fell between 1.0% and 3.5% above the true value. In this regard, it is interesting to note that the bias to the high side of component 12 declines with increasing degree of stretch, such that most errors in the region fell within the shaded histogram bin for the 12%-stretched condition. There is no reason to suppose that the change in absolute frequency across conditions is responsible for this reduction in bias. Instead, one might speculate that a progressive decoherence of the complex tone, caused by increasing the degree of stretch, results in component 12 becoming more like an isolated pure tone than like the upper spectral edge of a complex tone. This suggestion may merit further empirical study.

The other main feature of the error data is the growth across conditions of matches to the low- and mid-numbered components, which are better resolved (Plomp, 1976). However, it is not immediately clear why the growth in error matches to the low-numbered components is especially marked for component 3. Perhaps relevant here is the study of Moore and Ohgushi (1993), who used inharmonic complex tones with components that were equally spaced on an ERB (equivalent rectangular bandwidth) scale. The ability of listeners to hear out the interior partials from these complexes was often found to change in an irregular way across frequency. These authors suggested that this observation may reflect irregularities in the middle-ear transfer function, which alter the relative levels of components (Rosowski, 1991).

At least part of the difference in the pattern of errors between experiment 1 and 2 reflects the change in nominal $F0$ from 200 Hz to 100 Hz. This is clear from a comparison of the harmonic conditions in the two experiments. A reduction in matches to harmonic 1 might be expected when $F0$ is changed from 200 Hz to 100 Hz, because, at 60 dB SPL, a 100-Hz tone is 9.7 phons quieter than a 200-Hz tone (ISO 226, 1987). However, the other resolved harmonics will have changed in loudness by no more than 3–4 phons, and so the change in $F0$ should not of itself produce a general decline in matches to those components. The most likely explanation is the relationship between absolute frequency and the strength of the edge pitch evoked by component 12. Across conditions, the frequency of component 12 fell within the range 2160–2750 Hz in experiment 1 (nominal $F0 = 200\,\text{Hz} \pm 10\%$), and within the range 1200–2191 Hz in experiment 2 (nominal $F0 = 100\,\text{Hz} \pm 10\%$). As noted earlier, Hartmann *et al.* (1990) observed a progressive decline in the ability of listeners to hear out a mistuned harmonic for component frequencies greater than 2.2 kHz. Also, Kohlrausch and Houtsma (1991, 1992) reported that the pitch associated with the upper spectral edge of a harmonic complex was largely absent for component frequencies above 3 kHz.

IV. GENERAL DISCUSSION

A. Global pitch does not predict spectral fusion

The results of the experiments reported here are not consistent with the notion of a common mechanism that governs both the global pitch of a complex tone and the perceptual fusion of its components (see Hartmann, 1996). In particular,

the hit-rate profiles for experiment 1 do not vary with degree of frequency shift in a way that parallels changes in the value, strength, and ambiguity of global pitch across conditions. Evidence of a pitch/grouping dissociation of this kind has been provided by earlier studies (Ciocca, 1999; Brunstrom and Roberts, 2000, 2001).

Roberts and Brunstrom (1998) argued that the hit rates obtained for matching mistuned targets in 15%-shifted and 3%-stretched complex tones were too high to be accounted for simply by harmonic relations. The finding in experiment 1 that all degrees of frequency shift tested were well tolerated, and that the lowest overall hit rate was for the odd-harmonic stimuli (shift=50% of $F0$), supports the assertion that harmonic relations cannot account for the performance of listeners. Moreover, the lowest component of the 50%-shifted stimulus attracted many matches in error, despite its harmonic status. Nonetheless, it should be remembered that a proposal based solely on equal component spacing predicts that performance should not decline at all with increasing degrees of frequency shift.

The finding of experiment 2 that spectral stretch is effective at reducing matches to a mistuned target only when the degree of stretch is sufficient to produce short-range deviations from equal component spacing is consistent with Roberts and Brunstrom's (1998) proposal that the auditory system cannot utilize differences in component spacing between remote parts of the spectrum to fragment a complex tone perceptually. In contrast to the results of experiment 2, Roberts and Bailey (1996b) reported that a single even harmonic added to an otherwise odd-harmonic complex remained more salient than its odd neighbors after the application of a spectral stretch as great as 10%. However, this apparent inconsistency can be reconciled with the current findings by noting that the presence of the even-numbered component produces a local change in component spacing that dwarfs the consequences of the spectral stretch on the spacing of the other components. Therefore, it remains more salient.

Overall, the results indicate that Roberts and Brunstrom's (1998) proposal can account for the perceptual coherence of shifted and stretched stimuli, but only to a first approximation. This is perhaps unsurprising, given the simplicity of the proposal. An instructive approach to refining the proposal is to explore how stimuli of the kind used here might be represented in the auditory system. The following two sections describe the use of a contemporary model to simulate the auditory response to harmonic, shifted, and stretched stimuli, and consider how a central pattern recognizer might be able to pick out a mistuned component from these stimuli.

B. Representing auditory responses as autocorrelograms

Licklider (1951) proposed a ''duplex'' theory of global-pitch perception, in which a spectral analysis is accompanied by a periodicity analysis of the temporal fine structure of auditory nerve firings. His suggested mechanism was an autocorrelation analysis at each center frequency across the outputs of the auditory filterbank. Licklider's (1951) proposal has greatly influenced the development of computational models of auditory processing (e.g., Weintraub, 1985; Slaney and Lyon, 1990; Meddis and Hewitt, 1991a, b; Brown and Cooke, 1994a, b; Meddis and O'Mard, 1997). The Auditory Model Simulator (AMS), developed by Meddis and his colleagues, is a major contemporary model that is available to the research community as a working tool.[2] The simulations presented here were prepared using the AMS, and a list of all parameters used is given in the Appendix.

The AMS has not been altered substantially from the model originally proposed by Meddis and Hewitt (1991a). The reader is referred to their paper for full details. In brief, the model consists of four stages: (i) peripheral frequency selectivity (gammatone filterbank); (ii) within-channel half-wave rectification and low-pass filtering (hair-cell model); (iii) within-channel periodicity extraction (autocorrelation function, or ACF); and (iv) cross-channel aggregation of periodicity estimates (summary autocorrelation function, or SACF). A cross-channel set of ACFs can be depicted in a two-dimensional plot (frequency $\times$ delay period) called an autocorrelogram.

The core of the model is the use of the SACF to represent pitch information. Specifically, the most prominent peak in the SACF can be taken as providing an estimate of pitch. Meddis and Hewitt (1991a) used this approach to model successfully a number of key features of global-pitch perception (e.g., pitch of the missing fundamental, dominant region for pitch, and pitch shifts of frequency-shifted complexes). Meddis and Hewitt (1992) extended the model to tasks influenced by auditory grouping. They showed that human identification of concurrent vowels with different $F0$ frequencies can be approximated by: (i) using the main peak in the SACF to estimate the pitch of one vowel; (ii) selecting the set of channels with an ACF peak at this value as the auditory response to the corresponding vowel; (iii) selecting the remaining channels as the response to the other vowel; and (iv) performing a separate vowel-template match on each set of channels. Variants of this approach have since been incorporated into general computational models of auditory scene analysis (e.g., Brown and Cooke, 1994a, b).

Modelers have paid less attention to the auditory response evoked by a complex tone containing a single mistuned component. Meddis and O'Mard (1997) tried to model the influence of a mistuned harmonic on the pitch of an otherwise harmonic complex, but with only limited success. Data from human listeners indicate that global-pitch shifts are maximal for mistunings around $\pm 3\%$ (Moore *et al.*, 1985; Darwin *et al.*, 1994), but a model based on the best match of the SACF of a stimulus containing a mistuned harmonic to the SACFs of a set of truly harmonic comparison stimuli predicted maximal pitch shifts for mistunings twice this size. Meddis and O'Mard (1997) attributed this discrepancy to an additional mechanism leading to the perceptual segregation of the mistuned harmonic, but did not specify how their model could be modified to incorporate this factor.

We are unaware of any published examples of an autocorrelogram depicting the auditory response to a harmonic complex tone containing a single mistuned component. Part

(a) of Fig. 5 shows an autocorrelogram (top panel) and normalized SACF (bottom panel) for the harmonic stimulus from which our test stimuli were derived. Note the prominent peak in the SACF at a delay period of 5.0 ms, which corresponds to an $F0$ frequency of 200 Hz. The origin of this peak can be seen in the autocorrelogram, in which the alignment of common periods across frequency channels produces a characteristic vertical ''spine'' at the pitch period of the stimulus. Parts (b) and (c) differ from part (a) only in that component 4 is mistuned by -4% and $+4\%$, respectively. This component was chosen for mistuning because it is well resolved (Plomp, 1976) and falls in the middle of the dominant region for pitch (Ritsma, 1967). Note that the mistuned component leads to a ''ripple'' in the SACF, but that the most salient feature is a ''kink'' in the spine produced by channels responding primarily to the mistuned component. This may be the key feature in the model output that provides a basis for the perceptual segregation of the mistuned component.

C. Conjectures on a central pattern-recognition mechanism

Figures 6 and 7 show the equivalent autocorrelograms and SACFs for stimuli that are frequency shifted by 25% of $F0$ or spectrally stretched by 3% of $F0$, respectively. The nominal $F0$ was kept at 200 Hz in all cases to facilitate the comparison of auditory responses to harmonic, shifted, and stretched stimuli.[3] A comparison of Figs. 6(a) and 7(a) with Fig. 5(a) shows that shift and stretch cause a marked broadening of the main peak in the SACF, which is consistent with the weakening of pitch associated with these manipulations. The corresponding autocorrelograms show that the characteristic vertical spine seen in the auditory response to a harmonic stimulus becomes curved when the stimulus is shifted or stretched. Nonetheless, parts (b) and (c) of these figures show that mistuning component 4 still leads to a salient local perturbation in the spine. Therefore, we propose that a plausible account of the auditory system's sensitivity to mistuning in all three kinds of stimulus can be provided by the notion of a central pattern recognizer that operates on the whole autocorrelogram, rather than just on the SACF.

The decline in hit rates observed in the current study with increasing degrees of shift or stretch can be explained qualitatively by assuming that a kink in the spine of the autocorrelogram produced by a mistuned component becomes progressively more difficult to detect as the spine becomes more curved. Increasing the degree of either shift or stretch leads to a greater curvature of the spine away from the delay period of 5 ms in our exemplars, and this effect is especially marked over the range of stretches tested. This is consistent with the greater decline in hit rate observed across conditions in experiment 2 than in experiment 1. It is also noteworthy that the pattern of spine curvature differs for shifted and stretched stimuli (compare Figs. 6 and 7). For the shifted stimuli, the spine is most curved in the region of the autocorrelogram showing the auditory response to the low-numbered components. This is associated with a greater tendency for response errors to cluster around these components in the shifted conditions, suggesting that the low-numbered

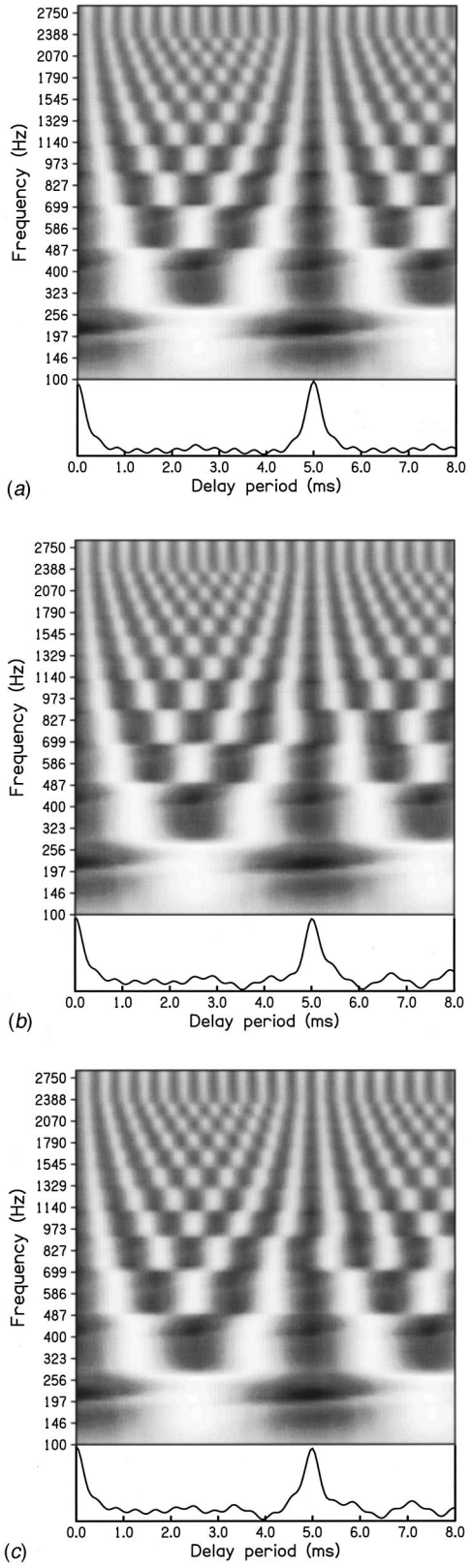
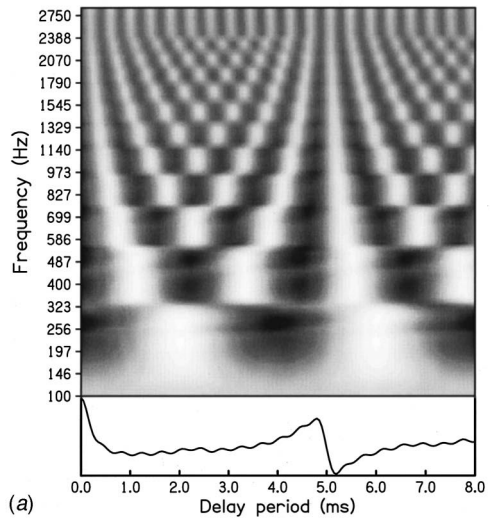
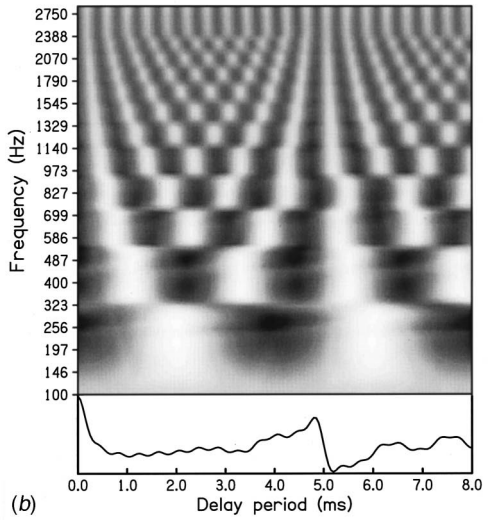


FIG. 5. Each part displays a set of cross-channel autocorrelation functions (an autocorrelogram) in the top panel, and a summary autocorrelation function (SACF) in the bottom panel, for a 12-component complex tone. Gray levels are used in the autocorrelogram to depict the cross-channel delay periods (higher values are shown using darker traces). Part (a) represents the model output for a harmonic complex tone with an $F0$ frequency of 200 Hz (component levels and phases are the same as those used in experiments 1 and 2). Parts (b) and (c) represent the model output when component 4 of the complex tone is mistuned by -4% and $+4\%$, respectively.





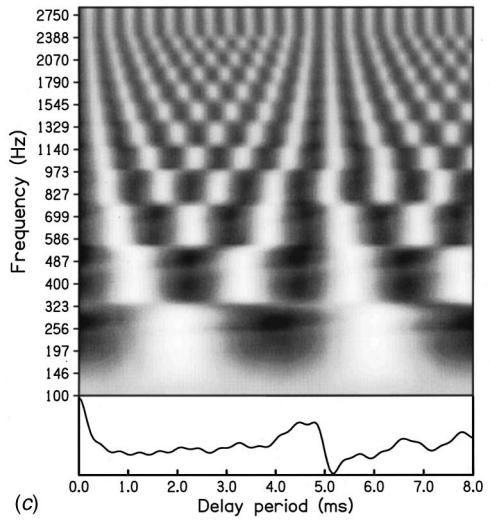
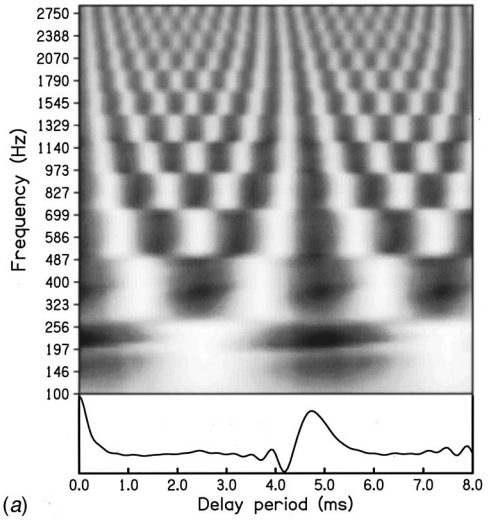
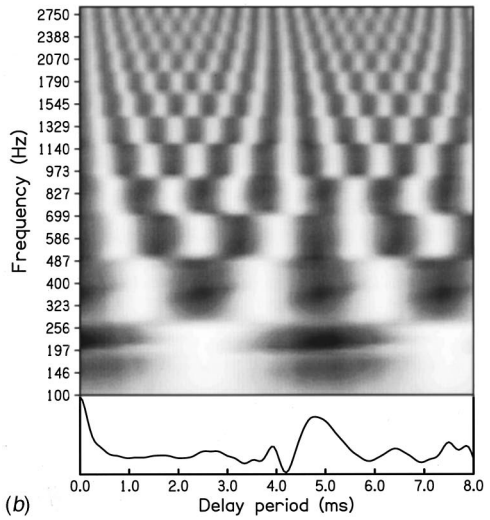


FIG. 6. Same as Fig. 5, except that the original complex tone is frequency-shifted by 25% of $F0$.





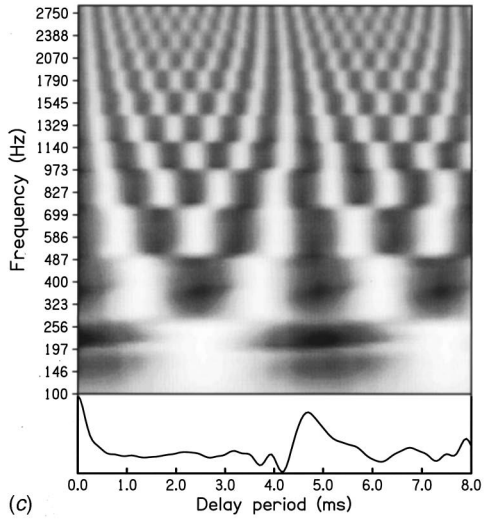


FIG. 7. Same as Fig. 5, except that the original complex tone is spectrally stretched by 3% of $F0$.

components of shifted stimuli are less well fused perceptually.

Our proposal about the role of the spine in the autocorrelogram is analogous to the weighted-image model of the effects of stimulus bandwidth on the lateralization of narrowband noises (Stern *et al.*, 1988; Trahiotis and Stern, 1989; Trahiotis *et al.*, 2001). Their model predicts the degree of perceived lateralization using a representation of the auditory

response as a cross-correlogram (frequency × interaural delay), and weights cross-frequency trajectories of maxima progressively less as these trajectories become more curved.

D. Concluding remarks

Combining information from both hits and misses can provide a useful insight into the perceptual fusion and fragmentation of complex tones. The idea that the auditory system is sensitive to deviations from equal component spacing, but only over a limited range, is broadly consistent with the current findings. Nonetheless, the observation that frequency shifting a harmonic complex produces a modest but significant decline in hit rate, and some peculiar features of the pattern of errors found for both shifted and stretched stimuli, indicate that Roberts and Brunstrom's (1998) proposal is too simplistic. The concomitant rise in incorrect matches to the low-numbered components with increasing frequency shift suggests that their role in the perceptual cohesion of a complex tone merits further investigation. In particular, it would be interesting to explore the effects on hit rate of mistuning the first component, which was not tested in our study.

An explanation based on features of the autocorrelogram can offer a more refined account of our findings than that provided by our original hypothesis. It may also offer further insight into the relationship between auditory grouping and the computation of global pitch. In particular, one might speculate that global-pitch perception relies on a cross-channel *aggregation* of periodicity information akin to the SACF (Meddis and Hewitt, 1991a, b; Meddis and O'Mard, 1997), but that the perceptual segregation of a mistuned component relies on a cross-channel *comparison* of periodicity information that looks for discontinuities in the otherwise smooth trajectory of an autocorrelogram spine. The latter suggestion is effectively a transposition of Roberts and Brunstrom's (1998) original hypothesis from the spectral structure of the stimulus to the spectrotemporal structure of the auditory response. While acknowledging that grouping is not an all-or-none process (e.g., Hukin and Darwin, 1995), we believe that grouping constraints of some kind must logically precede the process of cross-channel aggregation.

Developing the AMS so as to incorporate our proposed mechanism in a physiologically plausible way is a challenge for the future. Such an approach should also lead to a more successful simulation of the effects of a mistuned harmonic on global pitch than was achieved by Meddis and O'Mard (1997). There is also the question of what contribution might be made to perception by features in the autocorrelogram associated with higher-order delay periods. Other approaches to modeling the auditory response to shifted and stretched stimuli might also be explored. For example, Shamma and Klein (2000) have presented a radically new scheme whereby exposure to any broadband stimulus can lead to the generation of internal harmonic templates. The neural delay lines required by autocorrelation models are replaced in this scheme by a matrix of coincidence detectors. However, at this stage, it is not clear how Shamma and Klein's model could be modified to explain the observed perceptual coherence of shifted and stretched stimuli.

Whatever the details of the neural mechanism involved, it seems most unlikely that a set of predefined internal templates exists specifying various degrees of shift and stretch. More plausible is the idea that the auditory organization of these complex tones is a central but stimulus-driven phenomenon, evoked by features of the auditory response much like those represented in an autocorrelogram.

ACKNOWLEDGMENTS

This research was supported by Research Grant 6/S04782 from the Biotechnology and Biological Sciences Research Council (U.K.) to the first author. Special thanks go to Steve Holmes for his practical assistance with implementing the AMS, to Mike Harris for suggesting the use of luminance profiles to represent autocorrelograms, to Ray Meddis for discussions on the key features of autocorrelograms, and to Shihab Shamma for discussions on the possible nature of auditory templates. We are also grateful to Chris Darwin and an anonymous reviewer for their helpful suggestions on an earlier version of this manuscript.

APPENDIX: PARAMETERS USED FOR THE AUDITORY MODEL SIMULATIONS

Signal parameters	
1.00*E*-01	Stimulus signal duration (seconds)
5.00*E*-05	Stimulus sampling interval (seconds)
2.50*E*-03	Ramp-up rise time for signal (seconds)
Pre-emphasis filter	
2	Filter order
450	Lower 3-dB down cutoff frequency (Hz)
5000	Upper 3-dB down cutoff frequency (Hz)
Gammatone filter parameters	
4	Order for the gammatone filters
100	Lowest center frequency for filterbank (Hz)
2900	Highest center frequency for filterbank (Hz)
4.15	Equivalent rectangular bandwidth (ERB) rate
Hair cell parameters	
100.0	Permeability constant A (units per second)
6000.0	Permeability constant B (units per second)
2000.0	Release rate (units per second)
5.05	Replenishment rate (units per second)
2500.0	Loss rate (units per second)
66.31	Reprocessing rate (units per second)
6580.0	Recovery rate (units per second)
1.0	Maximum number of transmitter packets in free pool
50 000.0	Firing rate scalar (spikes per second)
Autocorrelation parameters	
8.0*E*-03	Maximum lag (seconds)
10.0*E*-03	Time constant (seconds)

[1]Chance performance can be estimated using the approach taken by Roberts and Brunstrom (1998). There are 12 consecutive components in each complex tone, giving a spacing of 11 units of component number between the lowest and highest partials. Given this, and the first criterion of $\pm 1.0 \times$ component-number units around the target frequency for a match to be classified as a hit, an estimate of 2/11, or about 18%, would seem a reasonable estimate. In practice, the use of a clustering algorithm as the second

criterion in the classification of responses may well have reduced the chance hit rate to an even lower value.

[2]The AMS is part of a computational library known as the Development System for Auditory Modeling (DSAM). The latest version can be obtained from the following web site: http://www.essex.ac.uk/psychology/hearinglab/

[3]Although the current study employed stretched stimuli with a nominal $F0$ of 100 Hz, the hit-rate data of Roberts and Brunstrom (1998) show that listeners are well able to segregate a mistuned fourth component in a 3%-stretched stimulus with a nominal $F0$ of 200 Hz. Therefore, it is legitimate to run simulations using stretched stimuli with this nominal $F0$.

Anderson, D. J. (**1973**). ''Quantitative model for the effects of stimulus frequency upon synchronization of auditory nerve discharges,'' J. Acoust. Soc. Am. **54**, 361–364.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).

Brown, G. J., and Cooke, M. (**1994a**). ''Computational auditory scene analysis,'' Comput. Speech Lang. **8**, 297–336.

Brown, G. J., and Cooke, M. (**1994b**). ''Perceptual grouping of musical sounds: A computational model,'' J. New Music Research **23**, 107–132.

Brunstrom, J. M., and Roberts, B. (**1998**). ''Profiling the perceptual suppression of partials in periodic complex tones: Further evidence for a harmonic template,'' J. Acoust. Soc. Am. **104**, 3511–3519.

Brunstrom, J. M., and Roberts, B. (**2000**). ''Separate mechanisms govern the selection of spectral components for perceptual fusion and for the computation of global pitch,'' J. Acoust. Soc. Am. **107**, 1566–1577.

Brunstrom, J. M., and Roberts, B. (**2001**). ''Effects of asynchrony and ear of presentation on the pitch of mistuned partials in harmonic and frequency-shifted complex tones,'' J. Acoust. Soc. Am. **110**, 391–401.

Ciocca, V. (**1999**). ''Evidence against an effect of grouping by spectral regularity on the perception of virtual pitch,'' J. Acoust. Soc. Am. **106**, 2746–2751.

Cohen, E. A. (**1980**). ''The Influence of Nonharmonic Partials on Tone Perception,'' Doctoral Thesis, Stanford University, CA.

Darwin, C. J., and Carlyon, R. P. (**1995**). ''Auditory grouping,'' in *Hearing: Handbook of Perception and Cognition*, 2nd ed., edited by B. C. J. Moore (Academic, London), pp. 387–424.

Darwin, C. J., Ciocca, V., and Sandell, G. J. (**1994**). ''Effects of frequency and amplitude modulation on the pitch of a complex tone with a mistuned harmonic,'' J. Acoust. Soc. Am. **95**, 2631–2636.

de Boer, E. (**1976**). ''On the 'residue' and auditory pitch perception,'' in *Handbook of Sensory Physiology, Vol. 5*, edited by W. D. Keidel and W. D. Neff (Springer-Verlag, Berlin), pp. 479–583.

Fine, P. A., and Moore, B. C. J. (**1993**). ''Frequency analysis and musical ability,'' Music Percept. **11**, 39–54.

Hartmann, W. M. (**1996**). ''Pitch, periodicity, and auditory organization,'' J. Acoust. Soc. Am. **100**, 3491–3502.

Hartmann, W. M., McAdams, S., and Smith, B. K. (**1990**). ''Hearing a mistuned harmonic in an otherwise periodic complex tone,'' J. Acoust. Soc. Am. **88**, 1712–1724.

Henke, W. L. (**1990**). MITSYN: *A coherent family of high-level languages for time signal processing*, software package (Belmont, MA).

Hukin, R. W., and Darwin, C. J. (**1995**). ''Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification,'' Percept. Psychophys. **57**, 191–196.

ISO 226 (**1987** (E)). ''Acoustics—Normal equal-loudness level contours,'' (International Organization for Standardization), pp. 1–7.

Kohlrausch, A., and Houtsma, A. J. M. (**1991**). ''Edge pitch of harmonic complex tones,'' IPO Annual Progress Rep. **26**, 39–49.

Kohlrausch, A., and Houtsma, A. J. M. (**1992**). ''Pitch related to spectral edges of broadband signals,'' in *Processing of Complex Sounds by the Auditory System*, edited by R. P. Carlyon, C. J. Darwin, and I. J. Russell (Oxford U.P., Oxford, U.K.), pp. 81–88.

Licklider, J. C. R. (**1951**). ''A duplex theory of pitch perception,'' Experientia **7**, 128–134.

Lin, J.-Y., and Hartmann, W. M. (**1998**). ''The pitch of a mistuned harmonic: Evidence for a template model,'' J. Acoust. Soc. Am. **103**, 2608–2617.

Mathews, M. V., and Pierce, J. R. (**1980**). ''Harmony and nonharmonic partials,'' J. Acoust. Soc. Am. **68**, 1252–1257.

Meddis, R., and Hewitt, M. J. (**1991a**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.

Meddis, R., and Hewitt, M. J. (**1991b**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity,'' J. Acoust. Soc. Am. **89**, 2883–2894.

Meddis, R., and Hewitt, M. J. (**1992**). ''Modeling the identification of concurrent vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **91**, 233–245.

Meddis, R., and O'Mard, L. (**1997**). ''A unitary model of pitch perception,'' J. Acoust. Soc. Am. **102**, 1811–1820.

Moore, B. C. J., and Glasberg, B. R. (**1988**). ''Effects of the relative phase of the components on the pitch discrimination of complex tones by subjects with unilateral and bilateral cochlear hearing impairments,'' in *Basic Issues in Hearing*, edited by H. Duifhuis, J. Horst, and H. Wit (Academic, London), pp. 421–430.

Moore, B. C. J., and Ohgushi, K. (**1993**). ''Audibility of partials in inharmonic complex tones,'' J. Acoust. Soc. Am. **93**, 452–461.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). ''Relative dominance of individual partials in determining the pitch of complex tones,'' J. Acoust. Soc. Am. **77**, 1853–1860.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). ''Thresholds for hearing mistuned partials as separate tones in harmonic complexes,'' J. Acoust. Soc. Am. **80**, 479–483.

Patterson, R. D. (**1973**). ''The effects of relative phase and the number of components on residue pitch,'' J. Acoust. Soc. Am. **53**, 1565–1572.

Patterson, R. D., and Wightman, F. L. (**1976**). ''Residue pitch as a function of component spacing,'' J. Acoust. Soc. Am. **59**, 1450–1459.

Plomp, R. (**1967**). ''Pitch of complex tones,'' J. Acoust. Soc. Am. **41**, 1526–1533.

Plomp, R. (**1976**). *Aspects of Tone Sensation* (Academic, New York).

Ritsma, R. J. (**1967**). ''Frequencies dominant in the perception of the pitch of complex sounds,'' J. Acoust. Soc. Am. **42**, 191–198.

Roberts, B. (**1998**). ''Effects of spectral pattern on the perceptual salience of partials in harmonic and frequency-shifted complex tones: A performance measure,'' J. Acoust. Soc. Am. **103**, 3588–3596.

Roberts, B., and Bailey, P. J. (**1996a**). ''Regularity of spectral pattern and its effects on the perceptual fusion of harmonics,'' Percept. Psychophys. **58**, 289–299.

Roberts, B., and Bailey, P. J. (**1996b**). ''Spectral regularity as a factor distinct from harmonic relations in auditory grouping,'' J. Exp. Psychol. **22**, 604–614.

Roberts, B., and Bregman, A. S. (**1991**). ''Effects of the pattern of spectral spacing on the perceptual fusion of harmonics,'' J. Acoust. Soc. Am. **90**, 3050–3060.

Roberts, B., and Brunstrom, J. M. (**1998**). ''Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes,'' J. Acoust. Soc. Am. **104**, 2326–2338.

Rosowski, J. J. (**1991**). ''The effects of external- and middle-ear filtering on auditory threshold and noise-induced hearing loss,'' J. Acoust. Soc. Am. **90**, 124–135.

Scheffers, M. T. M. (**1983**). ''Sifting Vowels: Auditory Pitch Analysis and Sound Segregation,'' Doctoral Thesis, University of Groningen, The Netherlands.

Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). ''Pitch of the residue,'' J. Acoust. Soc. Am. **34**, 1418–1424.

Shamma, S., and Klein, D. (**2000**). ''The case of the missing pitch templates: How harmonic templates emerge in the early auditory system,'' J. Acoust. Soc. Am. **107**, 2631–2644.

Slaney, M., and Lyon, R. F. (**1990**). ''A perceptual pitch detector,'' Proc. 1990 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), Albuquerque, NM, pp. 357–360.

Slaymaker, F. H. (**1970**). ''Chords from tones having stretched partials,'' J. Acoust. Soc. Am. **47**, 1569–1571.

Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (**1988**). ''Lateralization of complex binaural stimuli: A weighted-image model,'' J. Acoust. Soc. Am. **84**, 156–165.

Trahiotis, C., and Stern, R. M. (**1989**). ''Lateralization of bands of noise: Effects of band width and differences in interaural time and phase,'' J. Acoust. Soc. Am. **86**, 1285–1293.

Trahiotis, C., Bernstein, L. R., and Akeroyd, M. A. (**2001**). ''Manipulating the 'straightness' and 'curvature' of patterns of interaural cross correlation affects listeners' sensitivity to changes in interaural delay,'' J. Acoust. Soc. Am. **109**, 321–330.

Weintraub, M. (**1985**). ''A theory and computational model of auditory monaural sound separation,'' Doctoral Thesis, Stanford University, Stanford, CA.

Sources of variation in profile analysis. I. Individual differences and extended training

Ward R. Drennan[a] and Charles S. Watson
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47406

(Received 30 November 1999; revised 23 July 2001; accepted 3 August 2001)

This study investigated two sources of variance in the ability to discriminate auditory profiles: individual differences and extended training. The goals of the study were (1) to determine the range and origins of individual differences in profile analysis and (2) to determine whether those who initially had poor sensitivity to changes in spectral shape could eventually acquire finer sensitivity. Profile stimuli had 11 components with equal-log spacing from 200–2200 Hz. Thresholds ranged from −1 to −25 dB (signal level relative to the context level) across 46 listeners. The correlation between spectral-shape discrimination thresholds after 2000 trials and pure-tone intensity-discrimination thresholds was 0.36. The range of individual differences for pure-tone intensity discrimination and spectral-shape discrimination was about the same. Two groups of listeners were given extended practice on the profile task, one group that showed low thresholds after an initial 2000 trials of practice and another that showed much higher initial thresholds. All listeners improved during the course of the first 2000 trials of training. Most of the poor listeners continued to improve during 9000 trials of training. Individual differences in the listeners' sensitivity to changes in spectral shape still existed after the extended practice. [DOI: 10.1121/1.1408310]

PACS numbers: 43.66.Fe, 43.66.Jh, 43.66.Dc [MRL]

I. INTRODUCTION

''Profile analysis'' is a term used to describe listeners' detection of changes in the spectral shape, or profile, of a complex sound (Green, 1988). A spectral profile consists of a finite number of pure-tone components presented simultaneously. In profile analysis, the spectral shape is typically manipulated by adding an intensity increment to one component, thus creating a spectral ''bump.'' The ability to detect differences in spectral shapes is vital to understanding speech, appreciating music, and identifying environmental sounds.

This paper is the first of two describing a series of experiments conducted to determine and quantify the major sources of variance in profile analysis. Some of those sources of variance are a consequence of different abilities among listeners, while others appear to be associated with properties of the stimuli. These studies therefore employed an unusually large number of listeners and also investigated four different types of profile stimuli. The studies presented in this first paper address two important sources of variance that are apparent in the profile analysis literature: individual differences and extended training.

II. EXPERIMENT 1: INDIVIDUAL DIFFERENCES

A. Introduction

Substantial individual differences are evident in the profile analysis literature. Most studies have employed small numbers of listeners, and individual differences were sufficiently large that they sometimes had significant repercussions on the results of the experiments. The influence of individual differences was particularly apparent in the investigation of the effect of the number of tones on profile thresholds and also in the investigation of the profile bandwidth.

Green *et al.* (1983) and Green *et al.* (1984) showed that in profile analysis, energy outside the critical band of the signal *improves* performance. As the number of tones increased (and hence the bandwidth of the entire profile increased), sensitivity increased. Later studies showed that the magnitude of this effect was smaller than originally thought. Green *et al.* (1983) reported a 10-dB improvement as the number of components increased from 1 to 11, while Green (1992) reported a 2-dB improvement over the same range. Henn and Turner (1990) found a 2-dB improvement as the number of components increased from 3 to 11, and reported that the effect of increasing the number of tones was not universal, but highly individual.

Kidd (1993) addressed the issue of individual differences in profile analysis experiments in which the number of tones and bandwidth of the profile was increased. He repeated the experiments by Green *et al.* (1983) and by Green *et al.* (1984) and found substantial differences among 12 listeners. As the number of components was raised from 3 to 21, small groups showed improvements in threshold of 0 dB (four listeners), 6-8 dB (three listeners), 10 dB (three listeners), 14 dB (one listener), and 20 dB (one listener). Lentz and Richards (1998) measured the effect of increasing the number of components when the amplitude of the individual components was randomly selected, generating six different standard profiles. They found that there was no consistent effect of the number of components on profile thresholds for

[a]Presently at the Kresge Hearing Research Institute, Ann Arbor, MI, 48109-0506. Electronic mail: warddrennan@yahoo.com

four listeners. They suggested that the most likely reason for the discrepancy between these data and previous data was individual differences.

Neff and Dethlefs (1995) observed individual differences spanning a range of up to 59 dB for detection of a 1000 Hz tone amongst sinusoidal maskers whose frequencies were selected randomly on each presentation. While this study entailed a much higher degree of stimulus uncertainty than the other profile studies, it underscores the large variability that exists among listeners. A further analysis of the data including 49 listeners also revealed sex differences (Neff *et al.*, 1996). They observed average masked thresholds that were 7–8 dB higher for females than for males.

These studies demonstrate that individual differences exist in the profile analysis literature. They are sometimes large and can affect the results and conclusions of studies having a small number of listeners. The range of the differences, however, has not been established and only a modest amount of work (Berg *et al.*, 1993; Berg *et al.*, 1996) has investigated the origin of the differences. The present study aims to determine the range of and to further the study of the origins of individual differences in profile analysis.

B. Methods

1. Stimuli

The stimuli were 11-component profiles, digitally synthesized. The profiles had a duration of 400 ms including 5-ms, sine-squared rise ramps and cosine-squared fall ramps. All components were gated simultaneously in cosine phase. Static profiles were created with constant frequency over time and no added amplitude modulation. The components were spaced logarithmically in frequency starting at 200 Hz with a ratio of approximately 1.27 between adjacent components. This is referred to as a *static-log* profile, to distinguish it from other forms of profile stimuli utilized in later experiments in this series. The target tone was at the spectral center: 663 Hz.

Green (1988, p. 71) discussed the ''best'' standard profile which had log spacing greater than a critical band and equal-level components to minimize tone-on-tone masking. Other reports (Bernstein and Green, 1988) have shown that the minimum thresholds are obtained when the central frequencies of the profiles are incremented. The choice of stimulus parameters was based on these observations. The stimuli were presented using Tucker–Davis Technologies' (TDT) DA3 digital-to-analog converter. A 50-kHz sampling rate was employed. Level was controlled with a TDT SM3 module with a gain control for calibration. The overall level of the stimuli, presented binaurally to the listeners over TDH-39P headphones, was 70 dB SPL. The level of each individual component was 59.6 dB SPL.

2. Listeners

College students with normal hearing were recruited as listeners. All had hearing sensitivity of 15 dB HL or better from 250 Hz to 8 kHz. Forty-six listeners participated in the individual differences experiments (14 males, 32 females). Eleven of these listeners stayed on for the extended training experiment (five males, six females).

3. Procedure

This experiment was conducted in two parts. In the first part, 41 listeners were given the ''Test of Basic Auditory Capabilities'' (TBAC), developed by Watson *et al.* (1982). The TBAC test battery was included as a quick and reliable way to measure the listeners' performance on a variety of basic auditory tasks. This test battery examined eight auditory abilities using repeating series of six forced-choice trials. The trials within each series were ordered in increasing difficulty. The first three tests measured discrimination for increments in the frequency, level, and duration of single, 1000-Hz sinusoids presented at 75 dB SPL. The second three tests measured listeners' abilities to discriminate temporal changes in three types of complex sounds including pulse-trains, nine-tone patterns, and a sequence of four tones. The final two tests employed speech sounds. In one speech test, listeners were asked to discriminate a change in the sequence of four syllables. In the other, they identified nonsense syllables. Christopherson and Humes (1992) reported that the TBAC is a reliable test of auditory capabilities, noting that Cronbach's alpha was 0.79 or better for all the subtests except the speech identification tasks.

In the second part, the same 41 listeners, plus five who did not take the TBAC, detected increments in the level of the sixth sinusoidal component of 11-component profiles. The psychophysical procedure was a modified-2AFC method in which a standard profile was followed by two test profiles, one of which had an increment in the level of the middle component. Listeners were instructed to select the profile that sounded different from the standard. No special instructions were given. A light lasting 500 ms was illuminated 100 ms prior to the onset of each profile as a temporal cue. The interstimulus intervals were 500 ms. A two-down, one-up tracking procedure was used with 1-dB steps, converging on 70.7% correct (Levitt, 1971). Listeners were given feedback. Four listeners were tested together and all listeners had to respond before each successive trial was begun. Listeners typically responded within one second.

The experiment was conducted using 80-trial blocks. The final level visited in a block was saved and the next block began at that level. This procedure was used so that a continuous tracking history could be analyzed over the course of extended training. After each block, the mean level visited was reported to each listener. In each 90-minute daily session, the listeners completed 12 trial blocks. They received short (5–10 minute) breaks after each group of three blocks. The total number of trial blocks varied depending on the experiment. On average, each listener was tested for 2000 trials.

4. Method of threshold calculation

The tracking histories in this study consisted of about 500 reversals. Examination of these histories revealed that performance was asymptotic over the last 1/3 of the rever-

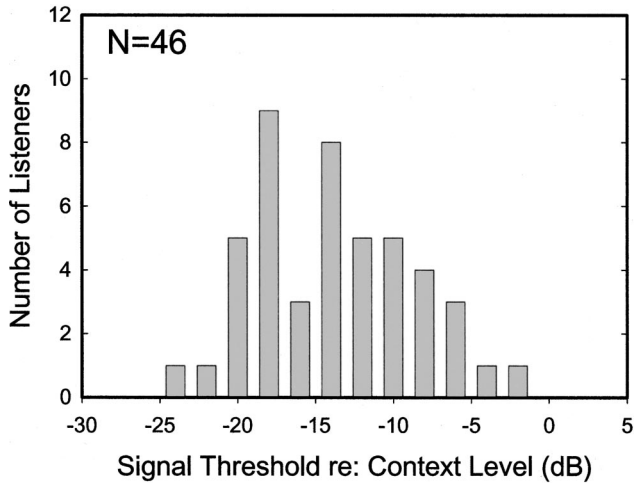


FIG. 1. Distribution of MMS threshold estimates for the static-log profile ($N=46$). See text for details.

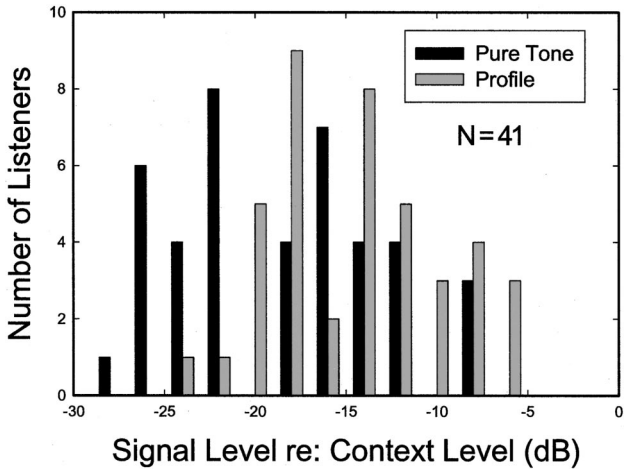


FIG. 2. Distributions of performance for static-log profile discrimination and pure-tone intensity discrimination for 41 listeners. Units have been transformed into ΔI in dB for both conditions.

sals, representing about 60 minutes of testing. In order to make reasonable use of the statistical power in this large quantity of data, five ''quasi-independent'' estimates of threshold were calculated for each listener and profile. The five threshold estimates were determined by calculating the mean of 10 reversals, skipping the next 20 reversals, calculating the mean of the next 10, and so forth. This was done in order to obtain a reasonable estimate of within-listener variance while preserving a stable threshold measure. For example, if 500 reversals were collected, the means of reversals 371–380, 401–410, 431–440, 461–470, and 491–500 were calculated. The overall threshold for a given stimulus condition was considered to be the average of the five means. In the remainder of this document, this method is referred to as the *mean of multiple samples* (MMS) method.[1] MMS thresholds with five samples encompass approximately 500 trials and can be calculated over any period of the tracking history. The thresholds reported here are calculated over the ending (asymptotic) period of the tracking histories unless stated otherwise.

C. Results

Figure 1 shows the distribution of thresholds calculated using the MMS method. Thresholds ranged from -0.96 to -24.66 dB (signal level relative to the context level[2]) with a mean of -13.98 dB and a standard deviation (SD) of 5.06 dB. Broken down by gender, 15 men averaged -15.40 dB with a SD of 4.44 dB, and 32 women averaged -13.23 dB with a SD of 5.25 dB. There was no significant difference between thresholds for men and women ($F_{1,44}=1.90$, $p=0.176$).

The TBAC scores were converted to the arcsine of percent correct. Correlations were calculated among these measures for each of the eight TBAC subtests and individual thresholds on the profile task. Three of the eight correlations among profile discrimination and the TBAC scores were significant at the 5% level: intensity discrimination (0.36), syllable identification (0.32), and embedded test-tone loudness (0.31). The embedded test-tone loudness test measures listeners' abilities to discriminate between a silent gap centered in a nine-component sequence and a 1.0 kHz tone presented instead of the gap, as a function of the gap duration. The test is labeled compatible with the listeners' experience; they are unaware of the existence of silent gaps in two of the three patterns presented on each trial. The overall score for the six nonspeech tests and the overall TBAC score (including speech) were also significantly correlated with the profile discrimination scores at the 5% level ($r=0.28$ and 0.29, respectively).

Figure 2 shows distributions of thresholds for the profile task and for the standard pure-tone intensity discrimination subtest from the TBAC battery. Among the subset of listeners ($N=41$) that participated in both profile and pure-tone intensity-discrimination tasks, the range of performance was -6 to -25 dB (signal level relative to the context level). A similar range of -8 to -29 dB was observed for pure-tone intensity discrimination. Measured as ΔI in dB, this corresponded to 0.3 to 3.0 dB for pure tones and 0.45 to 3.5 dB for profiles.

Table I shows the mean and standard deviation of these distributions using both measures. The pure-tone distribution clearly shows better performance than the profile distribution. A paired *t*-test showed that the difference between the means was significant ($t_{40}=3.52$, $p=0.001$). Discrimination ability was compared for a profile with an overall level of 70 dB SPL (59.6 dB per each component) and a pure tone of 75

TABLE I. Mean discrimination thresholds are shown with standard deviations for the two stimuli: profiles and pure tones. Two measures are used: signal level referenced to the context level as used in most profile experiments and ΔI in dB referenced to SPL. The ΔI in dB scale is much more compressive. (See Green, 1988, pp. 44–49.)

	Profile	Profile	Pure tone	Pure tone
Unit	Signal level *re*: context level	ΔI in dB	Signal level *re*: context level	ΔI in dB
Mean	-15.8	1.44	-18.9	1.09
Standard Dev.	4.3	0.68	5.56	0.66

dB SPL. No significant differences were observed in the variance of these distributions. This suggested that the individual differences observed in profile analysis were about the same magnitude as those for pure-tone intensity discrimination. The variance in sensitivity for pure-tone intensity discrimination was similar to that observed in other studies of individual differences (e.g., Johnson *et al.*, 1987 and Watson *et al.*, 1982).

D. Discussion

The ability to discriminate profiles was weakly but significantly correlated with the ability to discriminate differences in the intensity of a pure tone ($r=0.36$). It should be noted that this correlation relates intensity discrimination measured in a short time (72 trials on the intensity-discrimination subtest of the TBAC) and spectral-shape discrimination after 2000 trials of training. Pure-tone intensity discrimination was a simple task that required relatively little training, while profile analysis required up to 6 hours of training to approach asymptotic performance. Similar training time is needed in other complex sound discrimination tasks (Watson, 1991). The thresholds for pure-tone intensity discrimination, despite the fact that they were based on only 72 trials, were consistent with previous measures of intensity-discrimination capability (e.g., Berliner, 1973; Viemeister, 1972; Rabinowitz, 1970). The weak correlation between pure-tone intensity-discrimination and profile thresholds may reflect nothing more than similar central processing demands of these two tasks, either sensory or more generally cognitive. From these data alone, it is impossible to claim a common limitation originating in peripheral sensory mechanisms.

These data help explain an interaction observed by Green and Mason (1985). Green (1988, p. 96) expressed uncertainty as to the cause of the interaction. Green and Mason observed two groups of listeners. One group, experienced with spectral shape discrimination, had good sensitivity to spectral changes, but was not as good at pure-tone intensity discrimination. A second group, experienced with pure-tone intensity discrimination, had better sensitivity to pure-tone intensity than to spectral shape. Green stated that this was due to either a difference in training or a difference in the listeners. The inexperienced profile listeners were given 2000 additional trials of training, but little improvement was seen. This suggested that there might be two different types of listeners: one type particularly good at intensity discrimination that has difficulty with spectral-shape discrimination and another type that is good at spectral-shape discrimination but not so acute at pure-tone intensity discrimination. Thus, there may be an inverse relationship between pure-tone intensity discrimination ability and profile discrimination ability. Because the correlation between pure-tone intensity-discrimination and spectral-shape discrimination ability was decidedly positive, it appears that this was not supported. Green also wrote, ''It may be that differences in past experience can simply not be overcome by a few thousand trials.'' Based on the present data, it appears this interpretation was correct. The interaction observed by Green and Mason (1985) apparently was an effect of experience.

What could account for the individual differences? Intensity discrimination ability accounted for 13% of the variance among listeners. While this significant correlation between the two tasks was not surprising, the correlation was quite small. Green (1988, 1992) suggested that individual differences might also be attributed to differences in listening strategy, with poor listeners basing decisions on overall level and good listeners comparing levels simultaneously across channels. If overall level was used as a cue, a higher correlation may have been expected at least among the poor listeners. This was not observed in the data, suggesting that poor listeners do not base their decision on overall level. Studies by Berg and his colleagues suggested that poor listeners simply have poorer spectral resolution.

Berg *et al.* (1993) and Berg *et al.* (1996) defined a ''Central Processing Bandwidth'' (CPB) which, based on 50 listeners, could account for 27% of the variance observed in profile analysis. They conducted a band-widening experiment in which context tones, separated by 10, 20, 40, 80, or 160 Hz, were added to a profile stimulus. As more tones were added (increasing the bandwidth around the 1000-Hz target tone), sensitivity for the detection of an increment in the target tone decreased until a ''Fletcher-like'' break point was observed where the sensitivity began to increase. The width of the CPBs varied substantially across listeners, ranging from 1.5 to 4 times the Fletcher critical bandwidth. They observed a correlation of over 0.5 between the CPBs and profile[3] thresholds for the same listeners.

Berg and Green (1990, 1992) and Green and Berg (1991) determined the weight listeners placed on each component of a profile when making a decision. They observed that context tones close in frequency to the target component had higher weightings for poor listeners than for good listeners. This observation was consistent with Berg's CPB data.

Other empirical studies have suggested that individual differences in the peripheral auditory filters could also be significantly correlated with spectral-shape discrimination ability. In particular, data collected from hearing-impaired listeners who have wider than normal auditory bandwidths and higher thresholds for spectral-shape discrimination (Turner and Van Tasell, 1984; Leek *et al.*, 1987) suggested that normal listeners with wider auditory bandwidths may also have higher thresholds for spectral-shape discrimination.

III. EXPERIMENT 2: EXTENDED TRAINING

A. Introduction

Spiegel and Green (1982) wrote (p. 1209), ''How well a listener knows a particular spectrum depends on how often it is repeated.'' This provided an excellent summary of all the subsequent work on learning in spectral-shape discrimination studies. Over the course of several thousand trials on a single profile, thresholds approached asymptotic levels. Kidd *et al.* (1986) reported a mean improvement in profile thresholds of 10 dB from trial 300 to trial 3000.

Watson (1991) reported that discrimination tasks for complex sounds required 4 to 7 hours of training time for listeners to reach their best performance. The training time required for profile discrimination was consistent with the

TABLE II. MMS thresholds for good and poor profile listeners over four periods during the course of extended training. All MMS thresholds shown here are for the static-log profile. Listeners heard a total of 9000 trials (including 2000 trials of static-log, 2000 trials of static-harmonic, 2000 trials of dynamic-harmonic, 2000 trials of dynamic-log, and 1000 trials of static-log). The first column (Beginning) shows MMS thresholds calculated from reversals over the first 500 trials. The second column is MMS thresholds calculated from reversals over the last 500 trials of the first 2000 trials. The third and fourth columns show MMS thresholds calculated from reversals over the first and last 500 trials of the final 1000 trials of static-log profile discrimination. The final column shows the difference in threshold between the ending (asymptotic) portions of each static-log tracking history. Asterisks indicate significant differences. Standard deviations are shown in parentheses.

Listener	Beginning	End of first 2000 trials	Beginning of last 1000 trials	End of last 1000 trials	Difference between ends
Mean of good listeners	−11.3(3.0)	−20.8(3.0)	−19.3(4.7)	−20.0(3.5)	−0.8
L1	−15.7(9.3)	−25.5(1.1)	−23.2(2.5)	−23.2(2.0)	−2.3*
L2	−14.5(3.7)	−18.8(2.7)	−26.6(2.9)	−26.5(1.5)	7.7*
L3	−7.2(3.9)	−24.6(2.9)	−21.6(6.3)	−19.4(1.4)	−5.2*
L4	−9.3(3.7)	−19.2(3.2)	−17.5(4.2)	−17.9(1.7)	−1.3
L5	−9.3(4.7)	−18.6(2.1)	−14.4(4.4)	−17.9(2.8)	−0.7
L6	−11.2(5.3)	−19.3(3.6)	−15.7(6.2)	−16.9(1.2)	−2.4
L7	−11.9(6.4)	−19.3(1.9)	−15.9(1.0)	−18.6(1.8)	−0.7
Mean of poor listeners	−2.8(0.9)	−6.8(1.0)	−12.2(5.4)	−11.8(5.3)	5.0*
L8	−3.1(1.7)	−5.3(2.2)	−8.7(2.4)	−9.1(1.3)	3.8*
L9	−3.8(5.6)	−7.1(2.9)	−12.6(2.3)	−13.1(2.9)	6*
L10	−2.6(4.9)	−7.7(2.7)	−7.8(3.1)	−5.9(2.1)	−1.8
L11	−1.6(1.6)	−7.2(2.9)	−19.6(3.4)	−18.9(1.8)	11.7*

training time required for other discrimination tasks, such as frequency or duration discrimination (Watson and Kelly, 1978). More complex stimuli and more complex tasks, such as identification of speech stimuli, required much more training time; detection or discrimination with single tones required much less.

Green (1988, p. 94) suggested that in a profile-discrimination task, all listeners will improve to nearly the same level of performance, but some listeners are slow learners. For this reason, in Green's studies, listeners were screened and some listeners were dismissed because they were poor profile listeners initially. It has not been shown how much initially poor profile listeners actually improve with extended training. Thus, one question that has not yet been addressed is what are the individual differences in the abilities to learn to discriminate profiles? The experiment described below addresses this question.

B. Methods

The basic method for determining profile thresholds was the same as in Experiment 1. Seven ''good'' and four ''poor'' profile listeners were selected on the basis of their performance in Experiment 1 and on their availability. Good listeners had thresholds of −18 dB or better at the conclusion of experiment 1. Poor listeners had thresholds of −8 dB or worse after the same amount of training. (See Table II, ''End of first 2000 trials.'') Each listener heard at least 2000 trials of each of four profile types described and studied in detail in the companion paper (Drennan and Watson, 2001). The four types were (1) the static-log profile as in Experiment 1, (2) a static-harmonic profile with 11 harmonically spaced components from 200–2200 Hz, (3) a dynamic-log profile with log spacing and a rising frequency glide of 28% over 400 ms, and (4) a dynamic-harmonic profile with harmonic spacing and also a rising frequency glide of 28%. There were 2000 trials (2 days) of testing per profile type. All listeners began with the static-log profile and finished with an additional 1000 trials of the static-log profile. The order of presentation of the other profiles varied among listeners.

The initial purpose of the extended training study was to investigate the effects of training for good and poor listeners *and* to evaluate the difference among the four types of profiles. Unfortunately, confounding effects of the order of training were observed, so only the static-log data are presented. A more definitive study of the difference in sensitivity for different profile types is presented in the companion paper.

C. Results

Figure 3 shows a negative exponential curve of the form $y = k^* \log(1-\exp(-t/\tau)) + b$ fitted to raw tracking data for all individual listeners. *R*-squared values ranged from 50%–62% with a mean of 56% for good listeners and from 9%–31% with a mean of 19% for poor listeners. The size of the 95% confidence intervals of all asymptotes ranged from 0.4 dB to 4 dB. These estimates were about the same as the within-listener standard errors calculated using the MMS method with data from the end of the tracking histories. They ranged from 1 to 3 dB.

Table II shows MMS thresholds calculated from both the beginning and ending periods of tracking histories for the first 2000 and the last 1000 trials of static-log discrimination. The data showed that all the listeners improved during the initial period of training. Table II suggests that most (6 out of 7) of the good listeners did not continue to improve after the first 2000 trials of training (2 days). However, most of the poor listeners (3 out of 4) continued to improve during the 9000 trials of testing (9 days).

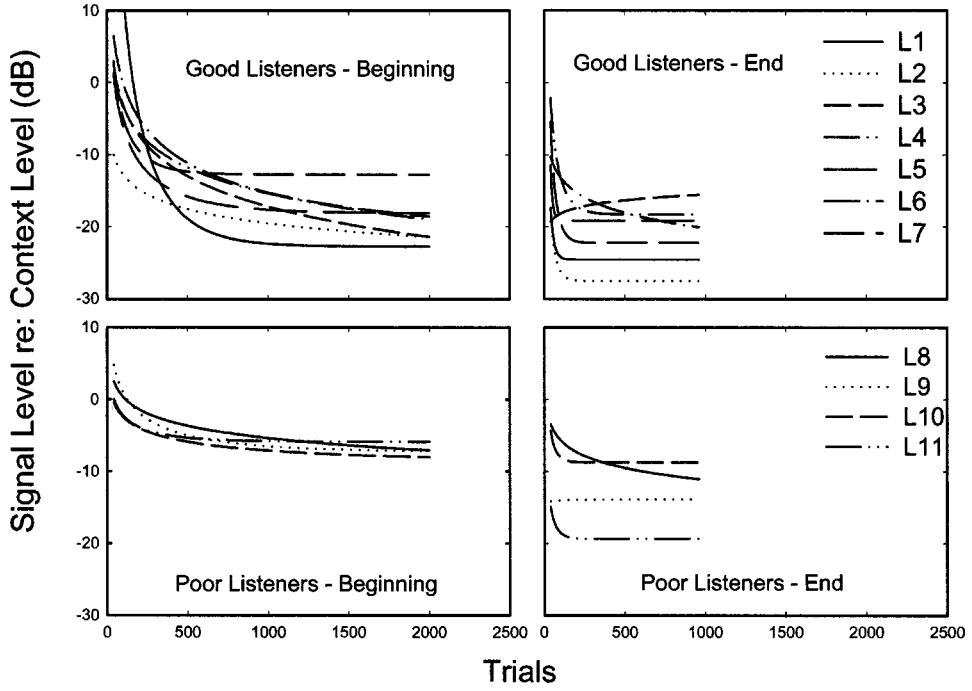


FIG. 3. Decaying exponential functions fitted to the tracking histories of static-log profiles heard by good and poor profile listeners at the beginning and end of a training period of 9000 trials (9 days).

These results demonstrated that learning rates differed among individuals. A large range of sensitivity remained after the training. One listener apparently could not learn to detect spectral changes despite prolonged training.

Note that these fits were calculated using raw tracking data. Given the tracking algorithm, these histories oscillate about a curve representing $d' = 0.76$ (70.7% correct). The *r*-squared values reflect these oscillations. Poor listeners had smaller *r*-squared values due to the fact that they exhibited little learning and perhaps because they had larger oscillations due to attentional fluctuation. The values of the asymptotes (a reasonable estimate of thresholds) were nevertheless always stable, with 95% confidence intervals smaller than 4 dB in every case.

D. Discussion

The present studies investigated profile analysis for both good and poor profile listeners. Many previous studies have only considered the performance of good listeners. It has sometimes been assumed that all listeners would eventually asymptote to nearly the same level. After 9000 trials of profile experience, listeners appeared to be approaching different asymptotic levels, and those levels differed by as much as 15–20 dB among listeners with normal pure-tone sensitivity. Three of the four poor listeners, however, did improve significantly during the course of 9000 trials of training. It could be argued that, given more training, these listeners would continue to improve, and therefore, for a majority of poor listeners, the assumption that they would improve to the level of the good listeners is correct. However, only one of the four poor listeners actually improved to the level of the good listeners. Furthermore, because the range of sensitivity remains quite large and is about the same magnitude as that observed for pure-tone intensity discrimination, it is perhaps unreasonable to expect that the sensitivity of listeners will converge on the narrower range of performance observed among the good listeners.

It should be noted that the listeners gained experience in the profile analysis task with the different types of profiles used in the companion paper. The original purpose of this extended training experiment was to investigate the performance on different types of profiles *and* the effects of extended training. Unfortunately, the order of training appeared to have an influence on performance. Because the profiles were not presented in a random order, the comparison among profile types was flawed; however, this still served as reasonable training. Had listeners been trained on the same profile throughout, the poor listeners may have improved even more. This supports the notion that most poor listeners may eventually reach the level of the good listeners. It should also be noted that the training with different profiles might account for the slight decrease in sensitivity observed in some of the good listeners.

Based on the present data, the validity of the assumption that poor listeners improve to the level of the good listeners cannot be determined. It can be concluded, however, that good listeners approach asymptotic levels within 2000 trials and poor listeners may take four or more times as long. Therefore, if the research goal is to determine the capabilities of the most sensitive listeners, and to do so in a short period of time, the practice of dismissing poor listeners in profile analysis studies (Green, 1988, p. 94) is justified.

ACKNOWLEDGMENTS

This work was supported by grants from the U.S. Air Force Office of Scientific Research and the NIH. Thanks to

Dennis McFadden, David Eddins, Marjorie Leek, Diane Kewley-Port, Gary Kidd, Donald Robinson, and an anonymous reviewer for providing helpful comments on earlier versions of this article. Thanks to Jay Inman for assistance with testing listeners. Thanks also to David Montgomery and Yijian Zheng for technical assistance.

[1]Thresholds calculated using the MMS method were highly correlated with thresholds determined by calculating the mean of the last 140 reversals for $N=46$ ($r=0.96$).

[2]The quantity ''dB (signal level relative to the context level)'' is equal to $20\log(\Delta A/A)$ in which A is the standard pressure of the context tones and ΔA is the increment in pressure on the target tone needed to detect a change in spectral shape. The formula for transformation from this measure to the difference limen in dB or ''ΔI in dB,'' more commonly used in intensity discrimination studies, is $x=20\log(10^{y/20}+1)$ where x is ΔI in dB. Note that ΔI in dB is a much more compressive scale. See Green (1988, pp. 44–49).

[3]Berg *et al.* used 5-tone, wide-band profiles (200–5000 Hz) with a 20 dB roving level.

Berg, B. G., and Green, D. M. (**1990**). ''Spectral weights in profile listening,'' J. Acoust. Soc. Am. **88**, 758–766.

Berg, B. G., and Green, D. M. (**1992**). ''Discrimination of complex spectra: spectral weights and performance efficiency,'' *Advances in Biosciences: Auditory Physiology and Perception 83* (Pergamon, Great Britain), pp. 373–379.

Berg, B. G., Southworth, C., and Turner, M. (**1993**). ''Critical bands for envelope cues,'' J. Acoust. Soc. Am. **93**, 2387(A).

Berg, B. G., Southworth, C., and Branstetter, B. K. (**1996**). ''Processing transition in a band-widening experiment,'' J. Acoust. Soc. Am. **99**, 2541–2542(A).

Berliner, J. E. (**1973**). Ph.D. thesis, MIT.

Bernstein, L. R., and Green, D. M. (**1988**). ''Detection of changes in spectral shape: uniform vs. non-uniform background spectra,'' Hear. Res. **32**, 157–166.

Christopherson, L. A., and Humes, L. E. (**1992**). ''Some psychometric properties of the test of basic auditory capabilities (TBAC),'' J. Speech Hear. Res. **35**, 929–935.

Drennan, W. R., and Watson, C. S. (**2001**). ''Sources of variation in profile analysis. II Component spacing, dynamic changes and roving level,'' J. Acoust. Soc. Am. , (in press).

Green, D. M. (**1988**). *Profile Analysis* (Oxford U.P., London).

Green, D. M. (**1992**). ''The number of components in profile analysis tasks,'' J. Acoust. Soc. Am. **91**, 1616–1623.

Green, D. M., and Berg, B. G. (**1991**). ''Spectral weights and the profile bowl,'' Q. J. Exp. Psychol. **43A**, 449–458.

Green, D. M., Kidd, Jr., G., and Picardi, M. C. (**1983**). ''Successive versus simultaneous comparison in auditory intensity discrimination,'' J. Acoust. Soc. Am. **73**, 639–643.

Green, D. M., and Mason, C. R. (**1985**). ''Auditory profile analysis: frequency, phase, and Weber's law,'' J. Acoust. Soc. Am. **77**, 1155–1161.

Green, D. M., Mason, C. R., and Kidd, Jr., G. (**1984**). ''Profile analysis: critical bands and duration,'' J. Acoust. Soc. Am. **75**, 1163–1167.

Henn, C. C., and Turner, C. W. (**1990**). ''Pure-tone increment detection in harmonic and inharmonic backgrounds,'' J. Acoust. Soc. Am. **88**, 126–131.

Johnson, D. M., Watson, C. S., and Jensen, J. K. (**1987**). ''Individual differences in auditory capabilities. I.'' J. Acoust. Soc. Am. **81**, 427–438.

Kidd, Jr., G. (**1993**). ''Individual differences in the improvement in spectral shape discrimination due to increasing number of nonsignal tones,'' J. Acoust. Soc. Am. **93**, 992–996.

Kidd, Jr., G., Mason, C. R., and Green, D. M. (**1986**). ''Auditory profile analysis for irregular sound spectra,'' J. Acoust. Soc. Am. **79**, 1045–1053.

Leek, M. R., Dorman, M. F., and Summerfield, Q. (**1987**). ''Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **81**, 148–154.

Lentz, J. J., and Richards, V. M. (**1998**). ''The effects of amplitude perturbation and increasing numbers of components in profile analysis,'' J. Acoust. Soc. Am. **103**, 535–541.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Neff, D. L., and Dethlefs, T. M. (**1995**). ''Individual differences in simultaneous masking with random-frequency, multicomponent maskers,'' J. Acoust. Soc. Am. **98**, 125–132.

Neff, D. L., Kessler, C. J., and Dethlefs, T. M. (**1996**). ''Sex differences in simultaneous masking with random-frequency maskers,'' J. Acoust. Soc. Am. **100**, 2547–2550.

Rabinowitz, W. M. (**1970**). M.S. thesis, MIT.

Spiegel, M. F., and Green, D. M. (**1982**). ''Signal and masker uncertainty with noise maskers of varying duration, bandwidth and center frequency,'' J. Acoust. Soc. Am. **71**, 1204–1210.

Turner, C. W., and Van Tasell, D. J. (**1984**). ''Sensorineural hearing loss and the discrimination of vowel-like stimuli,'' J. Acoust. Soc. Am. **75**, 562–565.

Viemeister, N. F. (**1972**). ''Intensity discrimination of pulsed sinusoids: The effects of filtered noise,'' J. Acoust. Soc. Am. **51**, 1265–1269.

Watson, C. S. (**1991**). ''Auditory perceptual learning and the cochlear implant,'' Ann. Otolaryng. Suppl. **12**, 73–79.

Watson, C. S., Johnson, D. M., Lehman, J. R., Kelly, W. J., and Jensen, J. K. (**1982**). ''An auditory discrimination test battery,'' J. Acoust. Soc. Am. Suppl. 1 **71**, S73.

Watson, C. S., and Kelly, W. J. (**1978**). ''Discrimination of tonal patterns with several levels of stimulus uncertainty,'' J. Acoust. Soc. Am. Suppl. 1 **64**, S39.

Sources of variation in profile analysis. II. Component spacing, dynamic changes, and roving level[a)]

Ward R. Drennan[b)] and Charles S. Watson
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47406

(Received 30 November 1999; revised 23 July 2001; accepted 3 August 2001)

Profile-analysis experiments have typically employed static profiles with constant frequency components spaced at equal intervals along a logarithmic frequency axis. Most periodic, naturally occurring stimuli, however, have components that are harmonically related and vary dynamically in time. One goal of these studies was to determine whether amplitude-increment detection thresholds are different in dynamic, harmonically spaced profiles compared to those for static-log profiles, and why such differences might exist. A second goal was to determine the impact of roving levels (within-trial variation of level). Thresholds for static-log profiles were, on average, 8.7 dB lower than for static-harmonic profiles. A traditional filter-bank model could not account for this result. No consistent effect of dynamic contour (an exponential rising frequency glide) was observed. Thresholds were consistently poorer by 4 to 7 dB when the level was roved, but the differences in thresholds among the different profiles varied little. It is proposed that the higher thresholds observed in static-harmonic profiles may be accounted for by the more intense pitch strength associated with the harmonic profiles.
[DOI: 10.1121/1.1408311]

PACS numbers: 43.66.Fe, 43.66.Jh, 43.66.Dc, 43.66.Mk [MRL]

I. INTRODUCTION

Profile analysis is the study of listeners' abilities to hear small changes in the spectral shape of a complex acoustic spectrum. Green (1988) provided an excellent review of early profile analysis studies. Profile stimuli consist of a number of equal-amplitude, static sinusoidal components spaced at equal intervals along a logarithmic frequency axis. Most naturally occurring periodic sounds, those generated by vibrating objects, however, have a harmonic (linear) relationship between components, and are often dynamic in frequency and amplitude. Few studies have investigated listeners' abilities to hear spectral changes in profiles with harmonic spacing (Henn and Turner, 1990; Zera *et al.*, 1993). No studies, to the authors' knowledge, have been published which investigate spectral-shape discrimination in profiles with dynamic pitch contours.

In most profile studies, the stimuli have been presented with the level randomly varied within trials. This is known as roving levels. While the roving-level technique reduces the utility of differences in overall level as the basis for discrimination, the additional stimulus uncertainty represented by within-trial level variation is also a cause of degraded discrimination performance. For this reason, the first experiments in this series (Drennan and Watson, 2001) did not use roving levels. In the present study, some conditions are conducted both with and without roving levels to provide a better understanding of its influence.

A. Component spacing

Differences in the spacing between components of a profile have had a substantial effect on the thresholds observed (Spiegel *et al.*, 1981). Kidd *et al.* (1986) determined thresholds for ten 11-component profiles with components positioned randomly along a frequency axis from 300–3000 Hz. The target was the center component of the profile. Listeners were tested for 1800 trials. Differences in thresholds for different profiles were not predicted well by the effects of local, energetic masking, although energy in the critical band around the target tone accounted for about 32% of the variance between average thresholds for the different profiles. Kidd *et al.* (1986) suggested that the energy distribution across the entire spectrum contributed to the thresholds. The results of Kidd *et al.* (1986) were consistent with those of Spiegel *et al.* (1981), Green *et al.* (1983), and Green *et al.* (1984), in that energy outside of the critical band had an effect on spectral-shape discrimination thresholds.

Green *et al.* (1983) studied the effects of component density. They observed that as long as context components were about 6% different in frequency from the target, thresholds decreased as the number of tones increased. This was true whether the number of tones was increased within a constant bandwidth or the spacing of the tones remained constant and the bandwidth increased. These results underscored the importance of context tones outside the critical band in profile analysis.

Henn and Turner (1990) raised another issue in regard to component spacing: is there a difference in profile thresholds between those with logarithmic and harmonic spacing? Henn and Turner investigated the differences due to component spacing, the effects of roving level, and the effect of number of components. In the logarithmic case, components had a frequency ratio of 1.38. In the harmonic case, the fundamen-

[a)]Portions of this study were presented at the third joint meeting of the Acoustical Society of America and the Acoustical Society of Japan, December 3, in Honolulu, Hawaii (Drennan and Watson, 1996).

[b)]Presently at the Kresge Hearing Research Institute, Ann Arbor, MI 48109-0506. Electronic mail: warddrennan@yahoo.com

tal was 200 Hz. In each case the target component was 1000 Hz. They found that, in a fixed-level condition using nine-component profiles, thresholds for the harmonic profile were about 8 dB higher than thresholds for the logarithmically spaced profile. This result could not be explained by the effects of local, energetic masking as assumed by traditional critical-band theory.

Using logarithmically spaced profiles, Bernstein and Green (1988) observed a U-shaped function when profile thresholds were plotted against signal frequency. When the signal was near the spectral edge of the profile, thresholds were about 10 dB higher than when the signal was at the spectral center of the profile. This result was inconsistent with critical-band theory and the effects of local masking.

Zera *et al.* (1993) studied harmonic profile discrimination extensively. They determined thresholds for harmonically spaced profiles with fundamental frequencies of 100, 200, and 400 Hz and 60, 30, and 15 components, respectively. Their results showed that differences between thresholds for log-spaced and harmonically spaced profiles depended on the frequency of the target component. Zera *et al.* explained the differences with traditional critical-band theory (energetic masking); however, this interpretation failed to account for Henn and Turner's report that harmonic profiles had much higher thresholds than log profiles.

B. Dynamic changes

The ability to identify auditory stimuli on the basis of different spectral shapes is essential for understanding speech, localizing sounds, and recognizing different timbres. Given that virtually any sound could be, in general terms, considered as having a dynamic spectrum, it is somewhat surprising that spectral-shape discrimination for dynamic spectra has not been studied more extensively. Static profiles were a logical place to begin in spectral-shape discrimination experiments, because they are simple in theory and easy to manipulate. The auditory system, however, evolved to detect naturally occurring sounds, most of which are dynamic and can be decomposed into harmonically spaced components. To date, only the effects of amplitude modulation in profiles have been studied in regard to dynamic spectral-shape discrimination (Green and Nguyen, 1988; Dai and Green, 1991; Bacon and Smith, 1996).

C. Roving level

Spiegel *et al.* (1981) and Green (1988) argued that listeners made a simultaneous comparison of level across multiple critical bands when discriminating spectral shape. To control for the possibility that listeners could base decisions on overall level, the researchers roved the level. Green reported that the effect of this manipulation was to increase thresholds 3 to 5 dB compared to discrimination without roving levels.

Mason *et al.* (1984) studied profile analysis and level variation in detail. In one experiment, they varied the range of level variation from 10 to 40 dB. Average thresholds for 10-, 20-, and 40-dB roving levels were −18, −14, and −13 dB, respectively (signal level relative to context level). In a second experiment, Mason *et al.* analyzed the effect of roving level by examining percent correct as a function of the difference in level between two successively presented profiles. The level of the profiles was varied between 20 and 80 dB SPL within trials (a 60-dB rove). Mason *et al.* (1984) showed that performance was best when the within-trial variation was less than 30 dB and became considerably worse (near chance levels) when the within-trial level difference was 60 dB.

Spiegel *et al.* (1981) provided arguments that roving level has little effect on how profiles are processed. They suggested that listeners used their frequency-resolving power in these experiments and did not base their decisions on a change in the overall loudness; therefore, a roving level should have little effect. In these experiments, the frequencies of the components of a profile were selected at random. These profiles varied in component spacing, but had a constant number of components; thus they had the same overall level. Spiegel *et al.* (1981) pointed out that the profile thresholds, if based on overall level, would be the same regardless of the number of components. The thresholds measured for the different profiles, however, were *not* the same. Thus, listeners must not have used overall level as a cue. Versfeld and Houtsma (1991) concurred, noting that the trend for thresholds with and without roving level was the same. Lentz and Richards (1997) showed that roving level increased uncertainty and raised thresholds, but had little effect on the observed differences in threshold *between* the different profile stimuli. Thus, as predicted by the model of Durlach *et al.* (1986), the rove simply added central noise, which was coherent across channels.

The following experiments were conducted to determine whether dynamic, harmonically spaced profiles yield different thresholds from the static-log profiles which have been traditionally used in profile experiments, and whether roving level has an impact on thresholds observed for the four types of profile stimuli described below.

II. METHODS

A. Stimuli

All profiles consisted of 11 components, were 400 ms in duration, and had 5-ms cosine-squared rise and fall ramps. The static-log profile had a ratio of approximately 1.27 between components covering a range of 200–2200 Hz. The static-harmonic profile had a fundamental of 200 Hz and harmonics up to 2200 Hz. The components of these static profiles remained constant in amplitude and frequency over the 400-ms duration of the stimulus. The dynamic profiles rose 28% in frequency over the duration of the stimulus. The temporal center of these profiles had the same instantaneous frequencies as the static profiles. The dynamic-log profile had a constant ratio of approximately 1.27 between components. The dynamic-harmonic profile had a continuous harmonic relationship between components. The beginning, middle, and ending frequencies of the dynamic profiles are listed in Table I. The target component in all cases was the sixth (middle) component (bold in the table). This component was selected as the target, because previous studies had

TABLE I. For 400-ms dynamic profiles, the starting, middle, and final frequencies of the harmonic and log profiles are listed. The middle frequencies indicate the frequencies of the static profiles. The frequencies in bold are the target frequencies. Values are rounded to the nearest Hertz.

Harmonic			Log		
Starting	Middle	Final	Starting	Middle	Final
175	200	224	175	200	224
351	400	449	223	254	285
526	600	673	283	323	362
702	800	899	360	411	460
877	1000	1123	458	522	586
1057	**1200**	**1353**	**582**	**663**	**745**
1228	1400	1572	740	843	947
1403	1600	1796	939	1071	1202
1579	1800	2021	1195	1361	1530
1754	2000	2245	1518	1730	1943
1930	2200	2470	1930	2200	2470

shown that minimum thresholds are obtained near the spectral center of the profiles (Bernstein and Green, 1988; Zera *et al.*, 1993). The frequency of the target component was 1200 Hz for the static-harmonic profiles and 663 Hz for the static-log profiles. While it would have been desirable to use the same target frequencies for the log- and harmonically-spaced profiles, that would have required either asymmetric spacing of the target for one profile or different frequency ranges for the log and harmonic profiles.

The stimuli are described analytically as

$$y(t)=\sum_{k=1}^{11}\cos(2\pi f_k t), \tag{1}$$

where the frequency of each component, k, is represented by f_k and the initial frequency of each component is represented by f_{ki}. In the static case, $f_k=f_{ki}$. In the dynamic case, $f_k(t)=f_{ki}e^{at}$ where $a=(\ln(f_{ki}/f_{kf}))/d$, f_{kf} is the final frequency, t is time, and d is the total duration of the stimulus in seconds. All components of the tonal complex were presented in cosine phase.

The stimuli were presented using Tucker–Davis Technologies' (TDT) DA3 digital-to-analog converter. A 50-kHz sampling rate was employed. Level was controlled with a TDT SM3 module with a gain control for calibration. A TDT PA4 programmable attenuater was used to control the amplitude in the roving-level conditions. The stimuli were presented binaurally to the listeners over TDH-39P headphones. The stimuli were presented at 70 dB SPL for the no-rove conditions (59.6 dB SPL per component). The roving conditions employed a 20-dB rove (± 10 dB) with a mean level of 60 dB SPL.

B. Listeners

College students aged 18 to 30 years were employed as listeners. All had audibility thresholds below 15 dB HL from 250 Hz to 8 kHz. A total of 43 naïve listeners were screened and seven listeners among the best half participated in the study.

C. Procedure

The four types of profiles (static-log, static-harmonic, dynamic-log, dynamic-harmonic) were presented in random order from trial to trial constrained by the requirement that no single type of profile would be heard twice in a row. This was done to eliminate any effect of the order of training. In an early study, the order of training appeared to be an important factor when evaluating listeners' abilities to discriminate spectral shape; therefore, it was considered necessary to eliminate the effects of the specific order of training by randomizing the order of stimulus presentation.

A two-down-one-up adaptive tracking procedure converging on 70.7% correct (Levitt, 1971) was used with 80-trial blocks. The levels visited were saved at the conclusion of each block such that the tracking histories continued for 2000 trials of each profile (8000 trials total). The last level visited in each block was reported to the listeners who also received trial-by-trial feedback. Listeners heard a standard profile followed by two alternatives, one of which contained an increment in the level of the middle component. Listeners were asked to select the profile that sounded different from the standard. The interval between stimuli was 500 ms. Thresholds were determined using the MMS method described in the companion paper (Drennan and Watson, 2001). This method calculated the means of the last 5 sets of 10 reversals, each separated by 20 reversals. Thus, for example, the mean of reversals 371–380, 401–410, 431–440, 461–470, and 491–500 would have been calculated in a tracking history with 500 reversals. This encompassed 500–600 trials. MMS thresholds were calculated to obtain 'quasi-independent' measures of thresholds used in estimating within-listener variance. Thresholds calculated in this way were highly correlated with thresholds calculated using the mean of the last 140 reversals ($r=0.96$).

In the profile literature, some controversy exists about the necessity of roving level. As argued by Spiegel *et al.* (1981) and Green (1988), it was possible that some listeners based their decisions on overall level rather than on spectral shape, making roving necessary. Later studies (Versfeld and Houtsma, 1991; Lentz and Richards, 1997), however, reported that roving level did not have any great effect, i.e.,

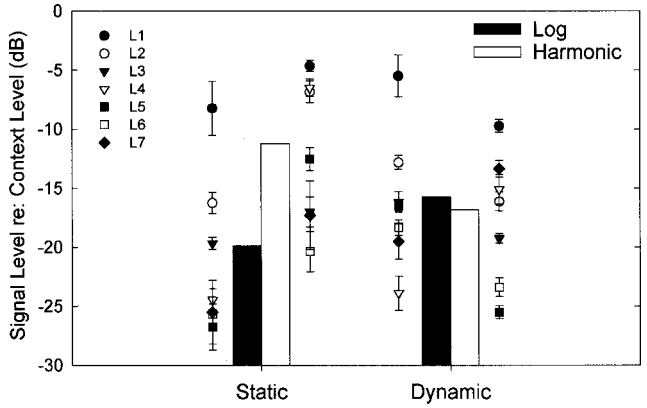


FIG. 1. MMS thresholds for static and dynamic profiles. The bars indicate the mean performance. The individual data are shown beside each bar. The error bars indicate the standard deviation observed with the five quasi-independent measures used in calculating the MMS thresholds.

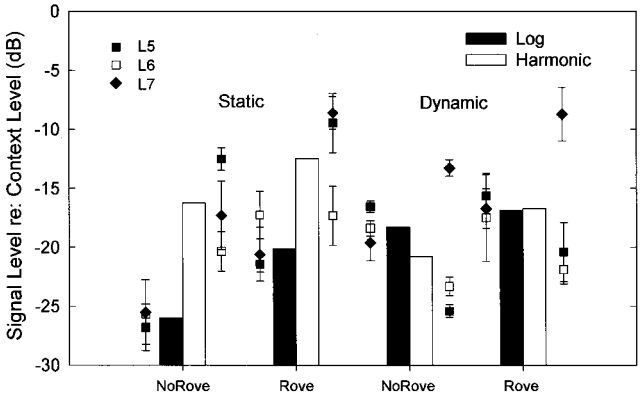


FIG. 2. MMS thresholds with and without roving level. The bars indicate the mean performance. Individual data are shown beside each bar. The error bars indicate standard deviation as in Fig. 1.

listeners may make decisions based on spectral shape even when an overall level cue is available, making roving level unnecessary. To determine the effects of within-trial level variation in the present study, one group of three listeners was tested with and without roving level.

The large group of 43 listeners was screened for one day (1000 trials) using the interleaved adaptive-tracking procedure with all four profiles. The seven best available listeners continued for seven additional days (8000 trials total including 2000 trials of each profile with interleaved tracking histories). Four of these listeners were tested only with fixed-level profiles. The remaining three listeners were first tested with roving level (8000 trials including 2000 trials of each profile) and then without roving level (4000 additional trials including 1000 trials of each profile).

III. RESULTS

Figure 1 shows MMS thresholds (signal level relative to context level) for seven listeners. Each symbol shows the data for a unique listener with error bars indicating standard deviations. The large vertical bars represent the means. All seven listeners had lower thresholds for the static-log profile than for the static-harmonic profile. Thresholds, on average, were 8.7 dB lower for static-log profiles than for static-harmonic profiles. This difference was significant: $t_6 = 4.104$, $p = 0.006$.

All of the data were analyzed with a three-way, repeated measures ANOVA [component spacing (log versus harmonic), temporal contour (static versus dynamic), and time of measure during the tracking history (5 sets of 10 reversals over the last 1/3 of the tracking history)]. No significant effect of spacing (log versus harmonic) was observed ($F_{1,6} = 4.892$, $p = 0.069$). No significant effect of temporal contour (dynamic versus static) was observed ($F_{1,6} = 0.733$, $p = 0.425$). There was no effect of the point in the tracking history in which data were collected ($F_{4,24} = 2.345$, $p = 0.083$) nor any interaction between that time and other factors. An interaction was found between spacing and contour ($F_{1,6} = 17.817$, $p = 0.006$.) This reflected the fact that five of seven listeners had lower thresholds for dynamic-harmonic than for dynamic-log profiles but higher thresholds for static-harmonic than for static-log profiles.

Figure 2 shows MMS thresholds for three listeners in the rove and no-rove conditions. Each symbol shows data for a unique listener with error bars showing the standard deviation. The large vertical bars represent the mean performance. The use of roving levels raised thresholds by 4–7 dB, but appeared to have no other substantial effect on the results. Other authors reported a similar result with different types of profiles (Versfeld and Houtsma, 1991 and Lentz and Richards, 1997).

IV. DISCUSSION

A. Component spacing

The most consistent result in these experiments was higher thresholds for the static-harmonic profiles than for the static-log profiles. Henn and Turner (1990) reported a similar difference. They showed that thresholds for 9-component profiles in their no-rove condition were 7–9 dB higher for static-harmonic profiles than for static-log profiles. There are a number of possible explanations for the difference between static-log and static-harmonic thresholds. The possible explanations are listed below. None of the individual causes could account for an average difference of 8 dB in the thresholds for static-log and static-harmonic profiles observed in the present experiments. It is possible that all of these factors, or some subset of them, combined to account for the differences observed.

(1) Energetic masking based on traditional models. A number of investigators have used excitation patterns to explain profile results (Spiegel *et al.*, 1981; Kidd *et al.*, 1986; Zera *et al.*, 1993; Ellermeier, 1996). A model based on the Moore and Glasberg (1987) excitation patterns is developed in the appendix. Applied to these data, the model predicts that the static-harmonic profiles yield thresholds 1 dB higher than static-log profiles. This difference is in the correct direction, but is smaller than the difference observed.

(2) Masking based on central processing bands. Based on the hypotheses set forth by Berg *et al.* (1993) and Berg *et al.* (1996), it is possible that listeners attempting to dis-

criminate profile differences have ''processing bandwidths'' which are 1.5 to 4 times the Fletcher critical bandwidth. If this were true, one would expect a greater amount of ''processing'' masking due to these wider bands. Narrow processing bands would yield the 1-dB difference as above. Extremely wide processing bands would result in high thresholds for both static-log and static-harmonic profiles. For the intermediate processing bandwidths, one would expect static-harmonic profiles to yield higher thresholds than the static-log profiles, because components of the static-harmonic profile were closer to the target than those of the static-log profile. Theoretically, the difference would be smaller for listeners with narrow processing bands. However, some of the most acute listeners in the static-log case exhibited the largest differences in thresholds. Unless the listeners employed different processing bandwidths for different profiles, it seems unlikely that variation in processing bandwidth could account for much of the difference observed.

(3) Absolute frequency of the target component. Another possible explanation is based on the physiological observations of Young and Sachs (1979) and Shivapuja *et al.* (1990). They showed that as the intensity of a single component increased, a spread of phase-locking occurred such that a wider band of neurons were phase-locked to the more intense components. If, as the data of Young and Sachs suggested, listeners could use this spread of phase-locking as a cue to discriminate spectral shape, then sensitivity to spectral shape would be worse for higher frequency targets, because the precision of phase-locking diminishes with increasing frequency. Thus, the difference in sensitivity observed between log-spaced and harmonically spaced profiles may be due to an effect of the absolute frequency of the targets.

Data shown in an article by Zera *et al.* (1993, p. 3435) suggested that at the frequencies tested in the present experiments, the harmonic profiles would have thresholds about 3 dB higher than the log profiles. This magnitude of the difference was smaller than the present study. This was probably due to the fact that their report compared thresholds from harmonic profiles in their own study to mean thresholds for log profiles from Bernstein and Green (1987).

(4) Effect of pitch strength. In the harmonic case, with the components in cosine phase, the intense pitch strength may have interfered with some listeners' ability to discriminate spectral shape. Gockel (1998) argued in support of this hypothesis, suggesting that pitch strength is a cue in profile analysis. In the present experiments, the pitch strength (closely related to the harmonicity of the components in cosine phase) seemed to dominate, at least subjectively, the perception of spectral shape. This phenomenon may be related to auditory grouping in which a higher-level construct, pitch strength, interferes with a lower-level analysis, spectral-shape discrimination. In a sense, this is similar to modulation detection interference (Yost *et al.*, 1989) in that a grouping phenomenon (in that case, modulation) interferes with detection. In the present case, the components are presumably grouped according to their harmonic relationships.

(5) Pitch as a cue. Another possibility is that pitch is a more effective cue in the log case than in the harmonic case. Dai (1993) noted that his variation on Feth's EWAIF model (SEWAIF) is most appropriate for profile stimuli. According to SEWAIF, no pitch cue is available in the static-harmonic case because the target is at the spectral center of the profile (on a linear scale). In the static-log case, SEWAIF predicts a change of pitch of 3.48 Hz for the average static-log threshold of −20.93 dB. According to SEWAIF, a pitch cue could enhance performance in the log case but not in the harmonic case. This explanation is weakened by the fact that the opposite result would be predicted if the ''center of mass'' were calculated on a log scale.

It also could be argued that the target component is closer to the region of dominance in the static-log profile than it is in the static-harmonic profile. Thus, an increment added to one component in the log profile would have a greater effect on pitch than an increment added to one component in the harmonic profile, consistent with lower thresholds for the log profiles.

(6) Temporal factors. The components for both profiles were in cosine-phase. In the harmonic case, this may have enhanced the salience of a temporal cue, because the wave form was quite peaked. Alcántara and Moore (1995) noted that such an increment in a component of a profile yielded a substantial increase in the amplitude of the wave form within the temporal valley. Thus, it would be expected that if listeners were using a temporal cue, performance would be better in the harmonic case with large temporal valleys. An increment would have a more substantial effect on the temporal envelope in the harmonic case than in the log case. The difference observed in this study was in the opposite direction. Even so, the temporal fluctuations may have interfered with the use of a pitch or other cue, thus raising thresholds in the harmonic case.

B. Dynamic contour

For dynamic profiles, data for individual listeners showed an inconsistent effect of component spacing. For most of the listeners (5 of 7), the effect was the opposite of that in the static case (dynamic-harmonic thresholds were lower than dynamic-log thresholds), although for two listeners the reverse was observed. Pitch strength may account for the differences in threshold observed between static-log and static-harmonic profiles, but these cues may have been obscured by the frequency glide in the dynamic profiles. It is possible that some type of ''pitch-strength integration'' takes place in the static case that could not occur in the dynamic case in which the pitch is constantly changing. If such an integration process differed sufficiently across listeners, it might yield the inconsistent results obtained here.

C. Roving level

The effect of roving level in these experiments was consistent with the data of Mason *et al.* (1984) for the static case. The 4–7 dB increase in average thresholds attributable to the roving levels in the present experiments was about the same difference as that observed by Mason *et al.* between a 10-dB rove and a 40-dB rove using a 21-component profile. Mason *et al.* did not report data for a no-rove condition.

The roving-level procedure was created to control for the use of overall level as a cue. If listeners used overall level as a cue, it would be expected that their sensitivity would be the same in every condition. Sensitivity was not the same, but was decidedly different for different profiles. Furthermore, roving level had little effect on the differences observed between profiles suggesting that it is unnecessary.

ACKNOWLEDGMENTS

This work was supported by grants from the U.S. Air Force Office of Scientific Research and the NIH. Thanks to Dennis McFadden, David Eddins, Diane Kewley-Port, Donald Robinson, Gary Kidd, and one anonymous reviewer for providing helpful comments on earlier versions of this article. Thanks also to David Montgomery for equipment maintenance and to Yijian Zheng for programming assistance.

APPENDIX

The excitation pattern model, based on the Moore and Glasberg (1987) excitation pattern, calculated a difference between the standard profile and the comparison profile in a ''noisy'' excitation pattern. This difference calculation was similar to the method used by Rice *et al.* (1995) to find the differences between spectra for sounds presented at different locations and to the method of Kewley-Port and Zheng (1998) who calculated the difference between vowel spectra generated by shifting formant frequencies. A version of the model was used to run simulations of tracking histories and to calculate thresholds based on the mean of several hundred reversals.

As input, the model was given the level and frequency of each component for the static-log or static-harmonic profile. The values were passed through a headphone transfer function to estimate the acoustic level as a function of frequency at the tympanic membrane (Bentler and Pavlovic, 1989). The information was then passed through the Moore and Glasberg excitation pattern program. As output, the Moore and Glasberg program generated an excitation pattern by calculating the level of excitation over each 0.1 equivalent rectangular bandwidth (ERB). The output contained 270 points across the acoustic spectrum.

Normally distributed noise equal to the quantity $m^*(1+\varepsilon)$ was generated independently for each 0.1 ERB and multiplied by the level of the excitation across the spectrum. The parameter ε was chosen randomly from a Gaussian distribution with mean$=0.0$ and variance$=1.0$. The parameter m was a fitting parameter chosen by hand so that the simulated tracking histories predicted thresholds near those observed by the best listeners (-25 dB signal level relative to the context level). The parameter m remained constant, and in this case was equal to 0.015. A larger m or variance for ε yielded higher threshold estimates while a lower m or variance yielded lower threshold estimates whether the input consisted of static-log or static-harmonic profiles.

The process of multiplying the output of the Moore and Glasberg excitation pattern by the random noise generated a ''noisy'' spectrum representative of the activity that might be observed in the auditory nerve. The process of generating this noisy spectrum was repeated for a standard pattern and two alternatives. The model then compared the standard and two alternatives by calculating the difference between two noisy excitation patterns and calculating the area under the difference curve. The alternative with the larger area was selected as the profile with an increment. A simulated tracking history was generated. Thresholds were calculated with the same MMS method used with the listeners' data. Twelve such simulations were run. The average threshold in the static-log condition was 1 dB less than the average threshold in the static-harmonic condition. This difference was in the same direction as that observed empirically; however, the magnitude was considerably smaller. Thus, this model based on the Moore and Glasberg excitation pattern could not account for the magnitude of the difference observed.

Alcántara, J. I., and Moore, B. C. J. (**1995**). ''The identification of vowel-like harmonic complexes: effect of component phase, level and fundamental frequency,'' J. Acoust. Soc. Am. **97**, 3813–3824.
Bacon, S. P., and Smith, M. A. (**1996**). ''Profile analysis with amplitude-modulated nontarget components,'' J. Acoust. Soc. Am. **99**, 1653–1659.
Bentler, R. A., and Pavlovic, C. V. (**1989**). ''Transfer functions and correction factors used in hearing aid evaluation and research,'' Ear Hear. **10**, 58–63.
Berg, B. G., Southworth, C., and Turner, M. (**1993**). ''Critical bands for envelope cues,'' J. Acoust. Soc. Am. **93**, 2387(A).
Berg, B. G., Southworth, C., and Branstetter, B. K. (**1996**). ''Processing transition n a band-widening experiment,'' J. Acoust. Soc. Am. **99**, 2541–2542(A).
Bernstein, L. R., and Green, D. M. (**1987**). ''The profile-analysis bandwidth,'' J. Acoust. Soc. Am. **81**, 1888–1895.
Bernstein, L. R., and Green, D. M. (**1988**). ''Detection of changes in spectral shape: uniform vs non-uniform background spectra,'' Hear. Res. **32**, 157–166.
Dai, H. (**1993**). ''On the pitch of two-tone complexes,'' J. Acoust. Soc. Am. **94**, 730–734.
Dai, H., and Green, D. M. (**1991**). ''Effect of amplitude modulation on profile detection,'' J. Acoust. Soc. Am. **90**, 836–845.
Drennan, W. R., and Watson, C. S. (**1996**). ''Discrimination of logarithmically and harmonically spaced static and dynamic spectral profiles with and without roving level,'' J. Acoust. Soc. Am. **100**, 2626(A).
Drennan, W. R., and Watson, C. (**2001**). ''Sources of variation in profile analysis. I. Individual differences and extended training,'' J. Acoust. Soc. Am. **110**, 2491–2497.
Durlach, N. I., Braida, L. D., and Ito, Y. (**1986**). ''Toward a model for discrimination of broadband signals,'' J. Acoust. Soc. Am. **80**, 63–72.
Ellermeier, W. (**1996**). ''Detectability of increments and decrements in spectral profiles,'' J. Acoust. Soc. Am. **99**, 3119–3125.
Gockel, H., (**1998**). ''On possible cues in profile analysis: Identification of the increment component,'' J. Acoust. Soc. Am. **103**, 542–552.
Green, D. M. (**1988**). *Profile Analysis* (Oxford U.P., London).
Green, D. M., Kidd, Jr., G., and Picardi, M. C. (**1983**). ''Successive versus simultaneous comparison in auditory intensity discrimination,'' J. Acoust. Soc. Am. **73**, 639–643.
Green, D. M., Mason, C. R., and Kidd, Jr., G. (**1984**). ''Profile analysis: critical bands and duration,'' J. Acoust. Soc. Am. **75**, 1163–1167.
Green, D. M., and Nguyen, Q. T. (**1988**). ''Profile analysis: detecting dynamic spectral changes,'' Hear. Res. **32**, 147–164.
Henn, C. C., and Turner, C. W. (**1990**). ''Pure-tone increment detection in harmonic and inharmonic backgrounds,'' J. Acoust. Soc. Am. **88**, 126–131.
Kewley-Port, D., and Zheng, Y. (**1998**). ''Auditory models of formant frequency discrimination for isolated vowels,'' J. Acoust. Soc. Am. **103**, 1654–1666.
Kidd, Jr., G., Mason, C. R., and Green, D. M. (**1986**). ''Auditory profile analysis for irregular sound spectra,'' J. Acoust. Soc. Am. **79**, 1045–1053.
Lentz, J. J., and Richards, V. M. (**1997**). ''Sensitivity to changes in overall level and spectral shape: an evaluation of a channel model,'' J. Acoust. Soc. Am. **101**, 3625–3635.

Levitt, H. (**1971**) ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Mason, C. R., Kidd, Jr., G., Hanna, T. E., and Green, D. M. (**1984**). ''Profile analysis and level variation,'' Hear. Res. **13**, 269–275.
Moore, B. C. J., and Glasberg, B. R. (**1987**). ''Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns,'' Hear. Res. **28**, 209–225.
Rice, J. J., Young, E. D., and Spirou, G. A. (**1995**). ''Auditory-nerve encoding of pinna-based spectral cues: rate representation of high-frequency stimuli,'' J. Acoust. Soc. Am. **97**, 1764–1776.
Shivapuja, B. G., Salvi, R. J., and Saunders, S. S. (**1990**). ''Response of auditory-nerve fibers to intensity increments in a multitone complex: neural correlates of profile analysis,'' J. Acoust. Soc. Am. **88**, 2211–2221.
Spiegel, M. F., Picardi, M. C., and Green, D. M. (**1981**). ''Signal and masker uncertainty in intensity discrimination,'' J. Acoust. Soc. Am. **70**, 1015–1019.
Versfeld, N. J., and Houtsma, A. J. M. (**1991**). ''Perception of spectral changes in multi-tone complexes,'' Q. J. Exp. Psychol. **43**, 459–479.
Yost, W. A., Sheft, S., and Opie, J., (**1989**), ''Modulation interference in detection and discrimination of amplitude modulation,'' J. Acoust. Soc. Am. **86**, 2138–2147.
Young, E. D., and Sachs, M. B. (**1979**). ''Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers,'' J. Acoust. Soc. Am. **66**, 1381–1403.
Zera, J., Onsan, Z. A., Nguyen, Z. T., and Green, D. M. (**1993**). ''Auditory profile analysis of harmonic signals,'' J. Acoust. Soc. Am. **93**, 3431–3441.

The intensity-difference limen for Gaussian-enveloped stimuli as a function of level: Tones and broadband noise[a)]

Lance Nizami,[b)] Jason F. Reimer, and Walt Jesteadt
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 3 January 2001; revised 10 August 2001; accepted 17 August 2001)

Van Schijndel *et al.* [J. Acoust. Soc. Am. **105**, 3425–3435 (1999)] have proposed that the internal excitation evoked by an auditory stimulus is segmented into ''windows'' according to the stimulus spectrum and stimulus length. This ''multiple looks'' model accounts for the mid-duration hump they observed in plots of intensity-difference limens (DLs) versus pip duration for Gaussian-shaped 1- and 4-kHz tones, an effect replicated by Baer *et al.* [J. Acoust. Soc. Am. **106**, 1907–1916 (1999)]. However, van Schijndel *et al.* and Baer *et al.* used few levels. A greater number of levels were used by Nizami (1999) for Gaussian-shaped 2-kHz tone-pips whose equivalent rectangular duration (D) was 1.25 ms. The DLs show the mid-level hump known for clicks [Raab and Taub, J. Acoust. Soc. Am. **46**, 965–968 (1969)]. At some duration this pattern must become the ''near-miss to Weber's law.'' To determine this duration, as well as the level-dependence of the mid-duration hump, DLs were established for Gaussian-shaped 2-kHz tone-pips of $D=1.25$, 2.51, and 10.03 ms at levels of 30–90 dB SPL. The across-subject average DLs for the tone-pips rise up at mid-levels for $D=1.25$ and $D=2.51$ ms. The DLs for $D=2.51$ ms are larger, creating the mid-duration hump. At all durations, the new DLs are smaller at high levels than at low levels, consistent with the near-miss to Weber's law. DLs were also obtained here for Gaussian-shaped broadband-noise pips of $D=0.63$, 1.25, 2.51, 5.02, and 10.03 ms. The DLs for the noise-pip show a mid-level hump for all pip durations. The noise-pip DLs decrease as the pip lengthens, such that the plot of DL versus log duration shows a linear decline, with no mid-duration hump. Analysis of variance reveals that the mid-level hump coexists with the classical patterns of level-dependence, perhaps reflecting the existence of two level-encoding mechanisms, one that depends on firing-rates counted over single neurons and which is responsible for the classical patterns, and one that depends on the initial coordinated burst of neuronal spikes caused by rapid ramping, and which presumably causes the mid-level hump. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1409371]

PACS numbers: 43.66.Fe, 43.66.Cb, 43.66.Ba [MRL]

I. INTRODUCTION

The ability to discriminate changes in the level of an auditory stimulus has been a focus of interest for many years, but little is known of the level-dependence of the intensity-difference limen (DL) for sinusoids of brief duration but relatively narrow bandwidth. Van Schijndel *et al.* (1999) have provided a theoretical rationale for establishing the discriminability of such stimuli. They assumed that the internal excitation evoked by an auditory stimulus is segmented into ''windows'' according to the stimulus spectrum and stimulus length, allowing multiple looks in the frequency domain and in the time domain. The largest DL would ensue for the stimulus expected to occupy the minimum effective area in the frequency-time space. They obtained DLs as a function of duration and sensation level for Gaussian-enveloped tone-pips of 6 durations ranging from 2.66 to 85 ms at 1 kHz, and 0.66 to 21.27 ms at 4 kHz, presented in a pink-noise background of 35 dB(A) SPL. The mean values of the DLs across three subjects revealed a notable mid-duration rise at 10 dB SL, but not at 0, 20, or 30 dB SL. Van Schijndel *et al.* found the poorest performance at durations of 4 ms at 1 kHz or 1 ms at 4 kHz, the durations assumed to represent a single frequency-time window.

Van Schijndel *et al.* (1999) used tone-pips whose amplitude envelope is Gaussian, because Gaussian envelopes yield the best trade-off between duration and bandwidth (Gabor, 1946), with the pip's energy appearing within a single, relatively narrow frequency lobe. For example, a 2-kHz tone-pip shaped with a Gaussian envelope of ± 4 standard deviations, comprising a total duration of 4 ms, has an energy-spectrum whose single lobe spans just 1601–2399 Hz at 6.8 dB below peak. Because the onset and offset ramps of a Gaussian envelope are gradual, the 4-ms tone-pip has an equivalent rectangular duration (D) of only 1.25 ms. Although integer total durations were used in generating the stimuli to be described later, the more transparent units of equivalent rectangular duration will be used in the remainder of this article.[1]

Baer *et al.* (1999) followed up on the van Schijndel *et al.* study using Gaussian-shaped tone-pips of 8 durations that spanned equivalent rectangular durations of 2–320 ms for 250-Hz tones, 0.50–80 ms for 1-kHz tones, and 0.10–20 ms for 4-kHz tones. DLs were obtained both with and with-

[a)]A portion of this work was presented in ''Intensity-difference limens as a function of level for brief 2-kHz Gaussian-shaped tones,'' 139th Meeting of the ASA, Atlanta, GA, June 2000. A separate portion was presented in ''Intensity-difference limens as a function of level for Gaussian-shaped packets of broadband noise,'' 140th Meeting of the ASA, Newport Beach, CA, December 2000.

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

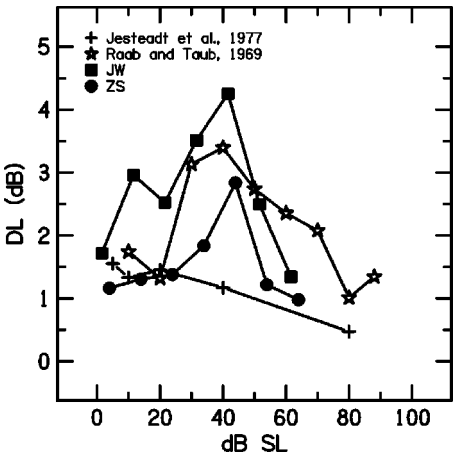


FIG. 1. Intensity-difference limens (DLs) for three kinds of stimuli. The solid squares and solid circles show the DLs for the Gaussian-shaped 2-kHz tone-pip for subjects JW and ZS, respectively. The open stars show the DLs for 0.1-ms acoustic clicks, averaged over two subjects (data of Raab and Taub, 1969). The plus signs show the DLs for 2-kHz tones of 500 ms averaged over three subjects (data of Jesteadt *et al.*, 1977).

out background noise. In the absence of noise, the tone-pips showed, on average, a mid-duration rise that peaked somewhere between $D=2$ and $D=5.02$ for the 1-kHz tone-pip, and that peaked at $D=2$ for the 4-kHz tone-pip.

Van Schijndel *et al.* (1999) and Baer *et al.* (1999) sampled only a few levels. Thus, the level-dependence of the DL is still not known for stimuli that are brief but relatively narrowband. There are a few exceptions; DLs have been reported as a function of level for the 2-kHz Gaussian-shaped tone-pip of $D=1.25$, for the condition in which each tone-pip was preceded by a weak forward-masker of the same frequency (Nizami and Schneider, 1994) and for the control condition, without a forward masker (Nizami, 1999). The DLs for the two subjects of Nizami (1999) appear in Fig. 1. For both subjects, the DLs peak at 70 dB SPL. Notably, listeners did not produce the ''near-miss to Weber's law'' (McGill and Goldberg, 1968), as seen, for example, for tones of 500-ms duration (Jesteadt *et al.*, 1977). Instead, the DLs resemble those found by Raab and Taub (1969) for 0.1-ms clicks, a pattern referred to henceforth as the ''mid-level hump.'' The Raab and Taub click DLs also resemble those known for brief, clicklike impulses of 0.3-ms duration (Avakyan and Radionova, 1963). The clicks and impulses share the characteristics of short duration and broad energy spectrum, whereas the 2-kHz 1.25-ms Gaussian-shaped tone-pips have an energy spectrum that is relatively narrowband. The agreement evident in Fig. 1 is thus intriguing. The present authors could find only one other study of the level-dependence of the DL for sinusoids of $D<10$ ms, that of Plack (1998), who obtained DLs for a 6-kHz stimulus having a 2-ms plateau duration and 2 ms onset- and offset-ramps. Plack found the same ''severe departure from Weber's law'' that had been noted by Carlyon and Moore (1984) for 6.5-kHz tones or 8-kHz tones of 26-ms steady-state duration and 5-ms onset- and offset-ramps, and replicated by Carlyon and Beveridge (1993) for 6.5-kHz tones of 25-ms steady-state duration and 2-ms onset- and offset-ramps. Such a pattern of level-dependence was not suspected to occur at 2 kHz. Carlyon and Moore (1984) saw no effect at 500 Hz and little if any at 4 kHz. At mid-frequencies, the mid-level hump was only known to occur under forward masking (Zeng *et al.*, 1991; Zeng and Turner, 1992; Plack *et al.*, 1995; Zeng and Shannon, 1995).

It should be noted that the level-dependence found by Nizami (1999) for Gaussian-shaped 2-kHz tone-pips of $D=1.25$ ms is not mandated by the mid-duration humps found by Baer *et al.* (1999) and van Schijndel *et al.* (1999) for their Gaussian-shaped tone-pips. Baer *et al.* (1999) and van Schijndel *et al.* (1999) made no strong assumptions about level, and might have obtained different results at higher levels. The duration-dependence of the DL that they reported begs the question of how the level-dependence would change with substantial alterations in the length of the Gaussian envelope. For 2-kHz tones, shorter pips might be expected to maintain the pattern already known for clicks (Fig. 1), but as the pip lengthens, some duration should be reached at which there is a transition from the clicklike pattern seen for $D=1.25$ ms, to the near-miss known for $D=500$ ms (Jesteadt *et al.*, 1977). It was suspected that this transition happens for durations closer to $D=1.25$ ms than to $D=500$ ms. Thus, DLs were obtained as a function of level for Gaussian-shaped tones of $D=1.25$, $D=2.51$, and $D=10.03$ ms. For simplicity's sake, no background noise was employed. Nine subjects were tested. Two of the nine subjects were also tested with tones of $D=0.63$ ms and of $D=5.02$ ms, and one of these subjects was further tested with tones of $D=3.76$ ms and $D=20.06$ ms.

The near-miss is generally attributed to spread of excitation as a function of level (e.g., Zwicker, 1970; Moore and Raab, 1974; Florentine and Buus, 1981; Viemeister, 1988). Increasing the bandwidth of low- and mid-level stimuli produces a comparable effect (Buus, 1990). Because the spectrum/duration trade-off obscures the possible role of duration in the level-dependence of the DL, a second experiment was run that was designed to minimize the effects of changing spectra, in which DLs were obtained using Gaussian-shaped packets of broadband noise (low-pass filtered at 8 kHz) having $D=0.63$–10.03 ms. The intensity DLs for these stimuli should be determined by duration alone, without a significant contribution from increased bandwidth at short durations. The multiple-windows model, therefore, predicts a monotonic increase in DLs at shorter durations. Interpretations of the noise-packet DLs must be tempered by the fact that poorer performance at shorter durations would be predicted on the basis of the statistical properties of the noise alone (e.g., Green, 1960).

II. METHOD

Two experiments were run. In the first experiment, DLs were obtained for Gaussian-shaped 2-kHz tone-pips. In the second, DLs were obtained for Gaussian-shaped packets of broadband noise. The same apparatus and procedure were used to establish the DLs in both experiments, and the subject pool was similar. Hence the subjects, apparatus, and procedure will each be described only once.

A. Subjects

Potential subjects were accepted only if their quiet thresholds for 200-ms sinusoids were ≤20 dB SPL at 500 Hz, 1 kHz, and 4 kHz in both the left and right ears. Quiet thresholds were obtained using the method described later in this work, with the exception that only a single block of trials was used to establish each threshold.

One listener was the first author (LN) and the rest were students from Creighton University. All but LN were paid volunteers, informed as to the method but not the expectations of the study. The average age of the first student group (tone experiment; two men, six women) was 22; the average age of the second student group (noise experiment; one man, four women) was 23. Subjects AE (male), MM (female), and LN (male) participated in both experiments. LN (age 40) had extensive listening experience with the 2-kHz Gaussian-shaped tone-pip of $D=1.25$ ms from a study that used the method of constant levels (Nizami and Schneider, 1999). The other subjects had no previous experience with any of these stimuli, although BB, RB, and PN had served in forward-masking experiments done in the BTNRH laboratory, experiments that utilized the same apparatus and similar procedures.

B. Apparatus and stimuli

All stimuli were generated digitally at a sampling rate of 50 kHz using an array processor (TDT AP2). Signals were played through 16-bit digital-to-analog converters (TDT DD1). The pedestal or comparison tone was generated on one channel of the DD1, beginning at a positive-going zero crossing, while the increment was generated in quadrature phase on the other. The output of each channel was low-pass filtered at 10 kHz (TDT FT6) and attenuated (TDT PA4), then the outputs of the two channels were combined (TDT SM3) and presented to the listener though a headphone buffer (TDT HB6), a remote passive attenuator in the soundproof chamber, and a Sennheiser HD 250 linear II headphone. Parallel use of multiple attenuators, summers, and headphone buffers made it possible to test as many as four listeners simultaneously. To take advantage of this arrangement, the subjects were assigned to groups if possible, and they remained within those groups. All stimuli were presented to the left ear. The level of the pip is reported as the level in dB SPL of the 2-kHz carrier. For measurements of quiet threshold, only the increment channel was used.

The stimuli were all created by multiplying carriers by Gaussian envelopes. In the first experiment, the carrier was a 2-kHz tone; in the second, it was a broadband noise. Each block of trials in the noise experiment started with a 10-s pause during which noise was generated to fill separate buffers for the pedestal and the increment, each buffer ten times longer than the required stimulus duration. The digital noise waveform was low-pass filtered using a FIR filter with a cutoff at 8 kHz. For each individual noise presentation, a segment of noise was cut from each buffer at a random starting point, and each such segment was then multiplied by a Gaussian envelope.

The Gaussian envelope was always smoothly reduced to zero amplitude at ±4 standard deviations (σ), so that stimulus duration was 8σ, where $8\sigma=8D/\sqrt{2\pi}\approx 3.19D$ (see footnote 1).

C. Procedure

Estimates of the intensity DL were obtained using a two-interval forced-choice adaptive-tracking procedure with a decision rule that estimates 70.7% correct (Levitt, 1971). Each adaptive track consisted of 50 trials, each trial consisting of two successive observation intervals. Both intervals contained the pedestal. Within each set of trials, the increment was randomly distributed between the first and second intervals, being replaced by zeroes (no stimulus) in the interval not containing the increment. Otherwise, the pedestal and the increment were of the same length, had the same Gaussian envelope, and appeared simultaneously whenever they occupied the same observation interval. The starting level of the increment was 15 dB higher than the level of the pedestal. For the initial trials, this starting level allowed the subjects to correctly differentiate between the interval containing the increment and that containing the pedestal alone. In each 50-trial block, 4-dB steps were followed by 2-dB steps after the fourth reversal. The reversal points for the 2-dB steps were averaged to obtain a threshold estimate, which is an increment level. As the increment was presented in quadrature phase to the pedestal, the difference limen is given by the equation $\mathrm{DL}=10\ \log_{10}\ [1+(\Delta I/I)]$ where $\Delta I/I = 10^{\mathrm{Threshold}/10}/10^{\mathrm{Pedestal\ level}/10}$. Learning effects were not evident, just as they were not evident for DLs obtained elsewhere using adaptive procedures [see Baer *et al.* (1999) for Gaussian pips, and Florentine *et al.* (1987) for longer tones]. Short-term learning was probably absorbed within the adaptive procedure itself.

The subjects observed a 16-character message window at the top of a keypad small enough to be held in the hands. At the start of each trial, a small asterisk appeared in the message window to indicate that stimuli were about to be presented. After 400 ms, the first stimulus interval started. The interval lasted 50 ms, with the stimuli starting at the beginning of the interval, followed by a 600-ms period of silence, followed by another 50-ms stimulus interval. (The stimulus interval was 100 ms for the Gaussian-shaped pips whose true total duration was 64 ms [$D=20.06$].) The subjects then had all the time they desired to choose the interval that they believed contained the increment, and to register their choice by pushing the button corresponding to that interval. Typically, no more than a few seconds was required. There was then a 500-ms delay before the next trial began. When more than one subject was tested at the same time, the 500-ms delay started at the time of the last response.

At the start of each of the experiments, the subjects were given three blocks of 50 trials as training: one block for a 30 dB SPL pedestal, one block for a 60 dB SPL pedestal, and one block for a 90 dB SPL pedestal. Subjects showed obvious proficiency at the task by the end of the third block. The first experiment, using only 2-kHz tones, was actually run in two halves. The first half involved subjects LN, SF, DW, SL, and RB. The second experiment was then run, using

TABLE I. Quiet thresholds for the Gaussian-shaped stimuli. Not tabled are the quiet thresholds for LN for the 12- and 64-ms tone-pips and the 12-ms noise-pip (18.31, 12.23, and 24.37 dB SPL, respectively).

	Duration D (ms)				
Subject	0.627	1.25	2.51	5.02	10.03
		Stimulus=2 kHz tone-pip			
BB		30.47	22.78		10.74
RB		27.22	22.41		16.51
AE		24.35	17.21		11.61
SF		26.05	21.39		16.84
SL		22.64	21.85		17.03
MM		24.16	22.46		16.67
LN	28.6	24.64	20.56	16.81	12.55
PN		30.92	27.77		16.91
DW	32.84	32.31	30.66	25.20	20.66
Mean±SD	30.72±2.12	26.97±3.27	23.01±3.73	21.01±4.2	15.50±3.01
		Stimulus=Noise-pip			
AE	28.01	24.36	21.3	15.89	14.66
BH	28.84	24.36	20.50	19.40	16.78
TH	30.41	30.33	26.48	23.23	20.02
MM	31.26	27.49	23.38	22.41	18.86
LN	31.7	30.57	24.96	23.82	22.14
RS	27.58	25.27	22.21	19.83	15.69
Mean±SD	29.63±1.58	27.06±2.61	23.14±2.07	20.76±2.73	18.03±2.58

Gaussian-shaped packets of broadband noise (described later in this work), and involved subjects LN, AE, MM, BH, TH, and RS. Subjects AE and MM volunteered for further testing, and so the first experiment was reopened, starting with three warm-up blocks. DLs were then obtained from AE and MM for 2-kHz tones, and DLs were also obtained for 2-kHz tones from two new subjects, BB and PN.

The subjects completed two 50-trial blocks at each pedestal level, moving from 30 to 90 dB SPL in steps of 10 dB. Two blocks used to measure quiet threshold were inserted into the sequence between the 40 and 50 dB SPL levels. This process was then repeated in reverse order, and then the whole cycle was repeated once more, providing a grand total of eight measurements of the DL at each stimulus condition. These DLs provided an arithmetic mean and a standard deviation, the latter used as a within-subject measure of variability.

DLs were obtained first for the 1.25-ms tone-pip, and then the above process was repeated for 10.03-ms tone-pips, and then for 2.51-ms tone-pips. Subject LN followed the same schedule as the others, but with the pedestal levels increasing in 5-dB increments from 30 to 90 dB SPL, not in 10-dB increments. DLs were obtained for the $D=1.25$, then $D=10.03$, and then $D=2.51$ conditions. DLs were subsequently obtained for $D=20.06$, 0.63, 5.02, and 3.76 ms, in that order. As a comparison to LN's results, DLs were also obtained from subject DW for $D=5.02$ ms, followed by $D=0.63$ ms.

In the second experiment, DLs were obtained for noise durations of $D=1.25$, 10.03, 2.51, 0.63, and 5.02 ms, in that order. For subject LN, pedestal levels changed in 5-dB increments, and DLs were also obtained for 3.76-ms pips.

III. RESULTS

A. Excluded data

The pedestal of 30 dB SPL turned out to be at or below the quiet threshold for some subjects. Hence DLs were included in the within-subject or across-subject results only when the subjects' quiet threshold was more than 5 dB below 30 dB SPL, a criterion based on the experimenter's interpretation of the subject's reports. Under this 5-dB criterion, four of nine subjects provided useful DLs at 30 dB SPL for the 1.25-ms 2-kHz tone-pip, and seven of nine subjects provided useful DLs at 30 dB SPL for the 2.51-ms 2-kHz tone-pip. None of the subjects provided useful DLs at 30 dB SPL for the 0.63-ms noise-pip, two of six subjects provided useful DLs at 30 dB SPL for the 1.25-ms noise-pip, and five of six subjects provided useful DLs at 30 dB SPL for the 2.51-ms noise-pip.

B. Quiet thresholds (experiments 1 and 2)

The quiet thresholds for the 2-kHz Gaussian-shaped tone-pip and for the Gaussian-shaped noise-pip appear in Table I. For the 2-kHz Gaussian-shaped tone-pip, straight lines were fitted to each individual subject's plot of quiet threshold versus duration using a nonweighted least-squares regression; the mean slope corresponded to -3.71 ± 1.27 dB per doubling.

The quiet thresholds of Abramovich (1978) show slopes of -2.51 dB/doubling for 2-kHz tones of 3 to 100-ms duration. Baer *et al.* (1999) found slopes of -3.2 and -2.7 dB per doubling of duration for 1 and 4 kHz, respectively (2 kHz was not used).

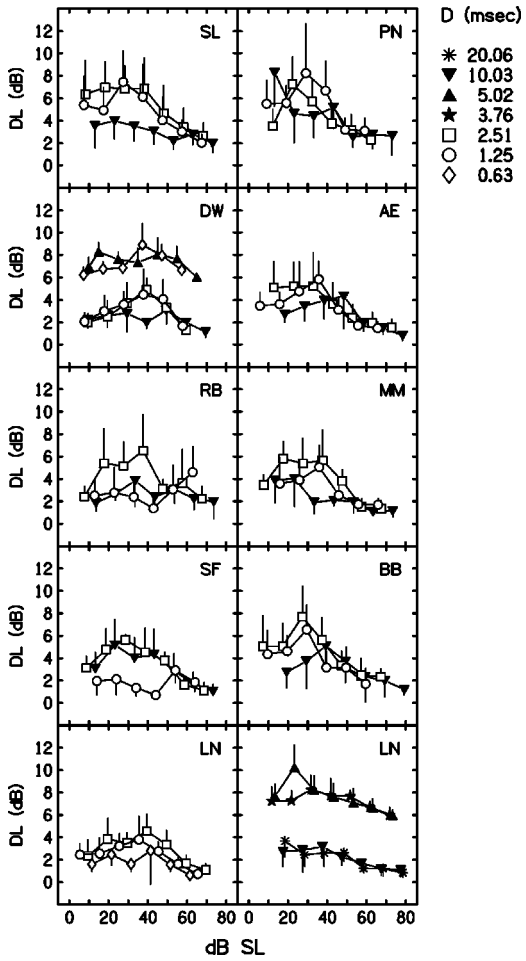


FIG. 2. The DLs for the 2-kHz Gaussian-shaped pip. Each panel represents an individual subject, and shows the mean DL and its standard deviation at each of three pip durations, $D=1.25$ ms (open circles), $D=2.51$ ms (open squares), and $D=10.03$ ms (inverted solid triangles). Also shown for subject DW are the mean DL and its standard deviation at $D=0.63$ ms (open diamonds) and $D=5.02$ ms (solid triangles). The latter two plots have been elevated by 5 dB for clarity. The bottom two panels show the mean DLs and their standard deviations for subject LN. In addition to DLs for $D=0.63$ and $D=5.02$, data for LN include DLs for $D=3.76$ ms (solid stars) and $D=20.06$ ms (asterisks). The latter two plots have been elevated by 5 dB for clarity.

The present noise-pip data were examined in the same manner as described earlier for the 2-kHz tone-pip. The mean of the individual subject's slopes corresponds to -2.95 ± 0.28 dB/doubling, which does not significantly differ from the -3.71 ± 1.27 dB/doubling found for the 2-kHz tone-pip (two-tailed *t*-test with unequal variances).

C. DLs for the 2-kHz tone (experiment 1)

1. Level-dependence of the DL

Figure 2 shows the DLs individually for the nine subjects. The error bars express the standard deviations across measurements. Clearly, the size of the DL and its pattern of level-dependence vary somewhat across subjects. Such individual differences are not unusual. They have been noted for longer tones (Florentine *et al.*, 1987), and despite such differences, meaningful patterns can be found in the average data (Florentine *et al.*, 1987). Figure 3 presents the mean DLs across subjects. To obtain the SL scale of the abscissa, the mean quiet threshold at each duration (Table I) was subtracted from each pedestal level. The DL for the 1.25-ms tone-pip rises at mid-levels, the peak apparently situated somewhere between 30 and 40 dB SL. The latter is in accord with the apparent peak at 40 dB SL for the DLs measured by Nizami (1999) for the tone-pip of $D=1.25$ ms (Fig. 1). Nizami's two subjects produced smaller DLs for pedestals at and below 40 dB SL, a difference most likely due to greater practice and less variability in stimuli from trial to trial. Figure 3 also reveals a mid-level rise for 2.51-ms tone-pips, with DLs that are larger than those for the 1.25-ms pip. SL scales are used because the peaks in the 1.25- and 2.51-ms plots appear to align in the SL scale, but did not align quite as well in an SPL scale.

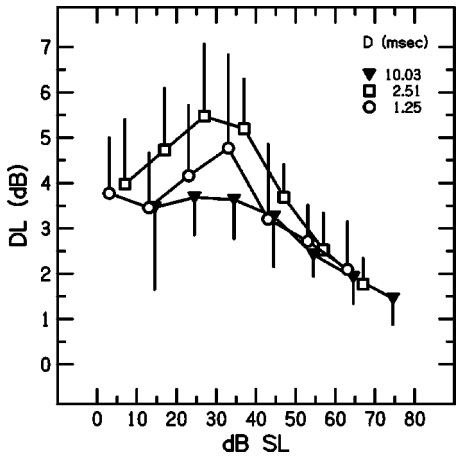


FIG. 3. The DLs for the 2-kHz Gaussian-shaped tone-pip, averaged for nine subjects. The error bars are the standard deviations across subjects. The abscissa is an ''average SL'' scale, constructed as described in the text.

The data in Fig. 3 were analyzed with a two-way (level by duration) ANOVA, using the interaction with subjects as the error term. The DLs for the lowest stimulus level (30 dB SPL) were not used, because not all subjects were able to provide DLs at that level (see above). For the level factor, conditions were grouped based on the SPL rather than the SL of the pedestal, because the same SPLs were employed at all durations. Significant effects were found for level [F (5, 40) $=25.594$, $p<0.001$], duration [F (2, 16)$=8.825$, $p<0.01$], and the interaction of level with duration [F (10, 80)$=2.517$, $p<0.05$]. In order to examine level effects in greater depth, a separate one-way ANOVA was completed for each duration. In this case, it made no difference whether the level categories were labeled in terms of SPL or SL, but SL values are used to refer to specific data points below. At each duration, the linear, quadratic, and higher-order components of the level-effect were evaluated (see Keppel, 1991, pp. 141–161). These components are orthogonal and provide a means of describing the nature of a significant level effect. The linear component represents a pattern in which the DL varies as a function of level but can be described by a straight line with nonzero slope, as in a ''near miss to Weber's law,'' while the quadratic component provides a measure of the significance of deviation from that linearity in the form of a single peak or trough, as in a mid-level hump. For the 1.25- and 2.51-ms tone-pips, linear components of the level effect were significant [at 1.25 ms, F (1, 8)$=7.663$, $p<0.05$; at 2.51 ms, F (1,

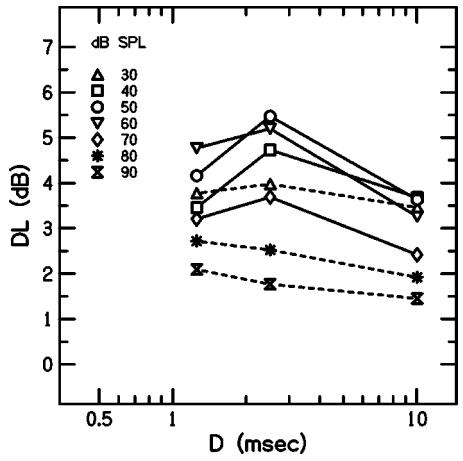


FIG. 4. The DLs for the 2-kHz Gaussian-shaped tone-pip, averaged over nine subjects (the data of Fig. 3), presented as a function of tone-pip duration.

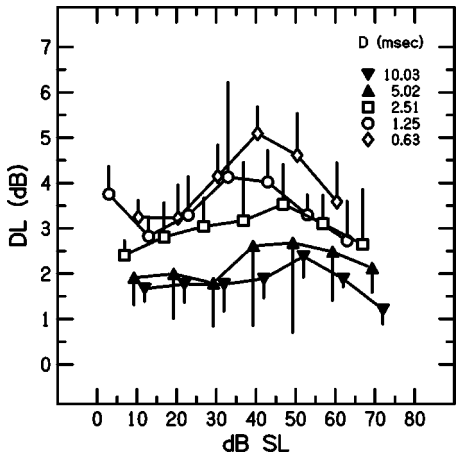


FIG. 5. The DLs for the Gaussian-shaped noise-pip, averaged over six subjects. The error bars are the standard deviations across subjects. The abscissa is an "average SL" scale, constructed as described in the text.

8)=46.823, $p<0.001$], quadratic components of the level effect were significant [at 1.25 ms, F (1, 8)=6.461, $p<0.05$; at 2.51 ms, F (1, 8)=20.762, $p<0.01$], and cubic components of the level effect were significant [at 1.25 ms, F (1, 8) =5.860, $p<0.05$; at 2.51 ms, F (1, 8)=6.727, $p<0.05$]. For the 10.03-ms condition, only the linear component was significant [F (1, 8)=54.015, $p<0.001$]. The DLs for the 10.03-ms tone-pip appear to follow the near-miss to Weber's law.

Next, pairwise comparisons were employed to test for significant differences between the visually identified peaks on each DL plot, and the points at levels above and below. For the 1.25-ms plot, the differences between the DL at 33 dB SL and the DLs at 13 and at all points above 33 dB SL were significant at $p<0.05$. The pattern is similar for the 2.51-ms plot, but the peak is not as well defined. For the 10.03-ms plot, the small rise evident in the DL occurs in the same region as it does for the shorter stimuli, but it is not significant. Once again, the DLs for the 10.03-ms tone-pip appear to follow the near-miss to Weber's law.

2. Duration-dependence of the DL

Separate one-way ANOVAs were carried out, as in the analyses described above, this time with pedestal level remaining constant for each ANOVA rather than stimulus duration remaining constant. The linear, quadratic, and other higher-order components of the duration effect were obtained at each pedestal level, to identify and evaluate any nonmonotonic trend. For each plot, pairwise comparisons were also employed. Figure 4 shows the mean DLs across all subjects, plotted as a function of tone-pip duration, with pedestal level as the parameter. Separate one-way ANOVAs were carried out for each plot. Significant linear effects were found for 80 dB SPL [F (1, 8)=16.047, $p<0.01$] and for 90 dB SPL [F (1, 8)=5.457, $p<0.05$] and significant quadratic effects were found for 40 dB SPL [F (1, 8)=6.429, $p<0.05$], 50 dB SPL [F (1, 8)=17.985, $p<0.01$], and 70 dB SPL [F (1, 8) =83.446, $p<0.05$]. The significant quadratic effects represent statistically significant mid-duration humps.

D. DLs for the broadband noise (experiment 2)

1. Level-dependence of the DL

The second experiment employed broadband noise as the carrier stimulus. For each subject and each pedestal level, the standard deviation across measurements generally increases in size as the size of the DL increases, as found for the 2-kHz tone-pip. However, the patterns of level-dependence for each subject were not as idiosyncratic as those found for the 2-kHz tone-pip (in Fig. 2), and so the individual DLs are not illustrated. Figure 5 shows the across-subject mean DLs, presented using SL scales that were constructed in the same manner as those used for the 2-kHz tone-pip. The DLs rise at mid-levels for pips of 0.63 ms, replicating the DLs for clicks (Raab and Taub, 1969), stimuli that have the same broad spectrum as do the noise-pips. A smaller mid-level rise is found for pips of 1.25-ms duration and a yet smaller rise for pips of 2.51-ms duration. The very large standard deviation for the 1.25-ms trace at about 30 dB SL is due to the unusually poor performance of just one subject, TH, at this level (pedestal=60 dB SPL).

The data of Fig. 5 were analyzed with a two-way (level by duration) ANOVA, using the interaction with subjects as the error term. For the level factor, conditions were grouped based on the SPL rather than the SL, as in the earlier analyses. Data for the lowest pedestal level shown in Fig. 5 were not included in the analysis because not all six subjects were tested at that level. Significant effects were found for level [F (5, 25)=9.316, $p<0.001$] and duration [F (4, 20)=19.363, $p<0.001$], as for the 2-kHz tone-pip, but significant effects were not found for the interaction of level with duration. The level effect has a highly significant quadratic component [F (1, 5)=21.194, $p<0.01$], but no significant linear component, suggesting that Weber's law (slope=0) holds at short durations for broadband noise stimuli, as it does for longer durations (Miller, 1947). Pairwise comparisons indicate a significant peak in all five plots of DL versus pedestal level. The peak was always located at 70 dB SPL. In Fig. 5 these pedestal levels correspond to 40, 43, 47, 49, and 52 dB SL, running from the shortest duration to the longest. The dura-

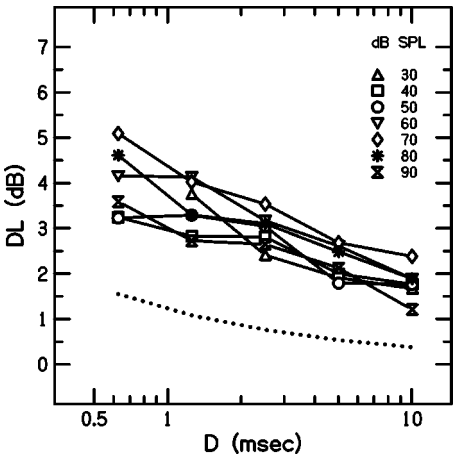


FIG. 6. The DLs for the Gaussian-shaped noise-pip, averaged over six subjects (the data of Fig. 5), presented as a function of noise-pip duration.

tion effect has a highly significant linear component [F (1, 5)=156.122, $p<0.001$], but no significant higher-order components.

2. Duration-dependence of the DL

In Fig. 6, the mean DLs are plotted versus noise-pip duration, with pedestal level as the parameter. Figure 6 reveals no mid-duration rise for the noise-pip, unlike the case for the 2-kHz pip (Fig. 4). The two-way ANOVA of these data showed a significant effect of duration, with no significant quadratic component, and a significant effect of level, but no interaction of duration with level. The level effect can be seen in Fig. 6 as the width of the band of DLs at each duration. Data for the 70 dB SPL conditions, the peaks of the mid-level humps, are at the top of the range.

On average, DLs for the noise pips in Figs. 5 and 6 are higher than for the tone pips in Figs. 3 and 4 and the change in DL as a function of duration is greater for the noise pips. The statistical properties of noise (de Boer, 1966; Green, 1960) might contribute to both of these effects. The variability in noise power from sample to sample is inversely related to WT, the product of bandwidth and duration, and the variability sets a limit on the size of the increment in noise that an ideal observer can detect. At small values of WT, this variability is sufficiently large relative to internal noise to limit performance. There is no corresponding limitation on performance for deterministic stimuli such as sinusoids, clicks, or frozen noise. For noise durations ranging from 3 to 300 ms and a noise bandwidth of 3862 Hz, Green found a constant 5-dB difference between the ideal-observer predictions and thresholds for human listeners when the predictions and thresholds were plotted as signal-to-noise ratios [$10\log(S_0/N_0)$]. Campbell (1963) found comparable but somewhat larger DLs over a range of durations from 1 to 100 ms for a noise bandwidth of 5000 Hz, and plotted the data both as signal-to-noise ratios and in the coordinates used here. The dotted line at the bottom of Fig. 6 shows predicted thresholds for the ideal observer for the present stimuli for $d'=0.78$, corresponding to 70.7% correct. In these coordinates, the difference between the performance of the ideal observer and the mean thresholds is somewhat larger at shorter durations. When the data are plotted in Green's coordinates, however, the difference is larger at the longer durations, ranging from 5.4 dB for the shortest stimuli to 7.4 dB for the longest stimuli. In either set of coordinates, the deviation between the ideal-observer predictions and the data changes monotonically with duration.

Performance of the ideal observer is best when the full bandwidth is used, but a human observer might well focus on a specific critical band. If it is assumed that W is 800 Hz, the approximate width of the critical band at 6 kHz, then the values of WT for the durations used here range from 0.5 to 8. Predicted thresholds for the ideal observer for these durations can be estimated from the psychometric functions for small WT given by Green and McGill (1970, Fig. 1). The values range from 6.9 dB at the shortest duration to 1.2 dB at the longest. Only the 6.9-dB value is outside the range of the mean data in Fig. 6. Thus an ideal observer using only a single critical band could still do as well or better than the performance represented by the present DLs for most, but not all, of the stimulus durations used here.

IV. DISCUSSION

A. General comments

One purpose of this study was to confirm that the intensity-difference limens for Gaussian-shaped 2-kHz tone-pips with a rectangular duration of 1.25 ms show a mid-level rise and then to explore at what durations that mid-level rise gives way to the near-miss to Weber's law. The second experiment employed a broadband-noise carrier in order to obtain results that might help to separate the possible contributions of stimulus duration and stimulus bandwidth.

The DLs for the tone-pip (Fig. 3) and those for the noise-pip (Fig. 5) all show a mid-level rise when the pips are sufficiently short, but the mid-level rise seen for the 2-kHz tone-pip is smaller for 1.25-ms duration than for 2.51-ms duration, creating a mid-duration rise in the DL. In contrast, the DLs for the noise-pip show an orderly increase in magnitude with decreasing length, consistent with the predictions of an ideal-observer model. This implies that the nonmonotonic duration-dependence of the DLs for the 2-kHz tone-pip is indeed an effect of bandwidth. The bandwidth of the shortest tone-pips ($D=0.63$ ms) is sufficiently broad [bandwidth $1/D$ at 6.82 dB down=1595 Hz, Baer *et al.* (1999), Eq. (2)] to result in improved performance relative to that obtained with a narrower bandwidth, but still sufficiently narrow that increases in stimulus level result in a near-miss to Weber's law. Duration itself seems to have two effects. The growth in the size of the DL is not very sensitive to duration; for tones of 250 Hz, 1 kHz, and 8 kHz with 1-ms ramps, shortening a tone causes a very slow increase in the DL (Florentine, 1986). But at some duration, the mechanism that causes the mid-level rise becomes notably dominant. The duration in question is greater than 2.51 ms, but is not greater than 10.03 ms for the 2-kHz tone-pip. Thus, the deviations from the underlying classical pattern become apparent only below some cut-off duration. For noise, the cut-off appears to be greater than 10.03 ms based on the mean data.

B. Implications for the multiple-windows model

Both the tone-pip DLs and the noise-pip DLs show a significant mid-level hump that is not predicted by the model of van Schijndel *et al.* (1999). The nonmonotonic level effect is in many ways a stronger feature in the data than the nonmonotonic duration effect; the level effect occurs for both tone and noise stimuli and the duration effect for tone stimuli occurs only at the mid levels. Support for the model comes from the fact that the DL for the broadband stimulus improves monotonically with duration, as the model would predict for a stimulus of constant bandwidth. As the statistical properties of the noise are a function of duration, it is of interest to ask whether use of frozen noise would have yielded different results. The most relevant data in the literature are for a comparison of level discrimination for random and frozen noise as a function of bandwidth (Buus, 1990). The analogous duration study has not been done. Over a range of WT values broader than those in the current study, Buus found a similar pattern of DLs as a function of WT for random and frozen noise with an equivalent rectangular duration of 27.5 ms. Of particular interest is the fact that the DLs were poorer at the mid level, 60 dB, than at lower and higher levels, 30 and 90 dB, for both random and frozen noises at all bandwidths. If varying T is equivalent to varying W, these results suggest that DLs would be poorer at mid-levels for frozen as well as random noise at all durations. Thus although the statistical properties of the noise may have contributed to the increase of the DLs as duration shortens (Fig. 6), these properties probably did not determine the shapes of the curves seen in Figs. 5 and 6.

The conclusions are similar if we consider the statistical properties of the noise that falls within a given critical band. As mentioned in the section on the time-dependence of the DL for the Gaussian noise-pip, performance at the shortest duration used here is better than the ideal-observer model would predict for a critical band at the upper end of the available frequency range, suggesting that the listener uses information from multiple windows. Increasing the pedestal level might widen the effective critical band, both because higher frequencies would be audible, and because the width of the critical band might increase with level. A wider critical band would allow better performance for an energy detector focused on that critical band. These effects would be small, however, and monotonic with level, unlike the results shown in Fig. 5.

C. A physiological model of the mid-level hump

The results shown in Fig. 3 confirm the Nizami (1999) observation of a mid-level hump in the DLs for the 2-kHz Gaussian-shaped tone-pip (Fig. 1). The level-dependence is reminiscent of the just-detectable intensity-increments (Raab and Taub, 1969) or the just-detectable intensity-decrements (Avakyan and Radionova, 1963) known for clicks. The hump observed for clicks was explained by Taub and Raab (1969) in terms of the variability of the click-evoked N_1 component of the compound action potential (CAP). Taub and Raab first obtained neurometric DLs in the Guinea pig, in much the same way that psychophysical DLs have been established in humans. The resulting DLs peaked at about 38–48 dB relative to a human listener's click-detection threshold. Taub and Raab (1969) also found that amplitude histograms for click-evoked N_1 were unimodal for the lower ranges of click intensity, and approximately Gaussian. Standard deviations of the N_1 distribution were estimated over similar intensity ranges for three animals, giving six to seven data points for each animal. These standard deviations rose with level and then started to decline, showing a maximum within the region of maximal Weber fractions for humans (Raab and Taub, 1969). Raab and Taub concluded that the variability in N_1, represented by its standard deviation, accounted for the observed mid-level rise in human DLs for clicks (Raab and Taub, 1969).

This conclusion suggests that variability in the N_1 might also account for the mid-level rise seen in Figs. 3 and 5. Human subjects do indeed produce an N_1 that can be visually distinguished from the noise floor, and whose amplitude generally increases as a function of stimulus level, in response to Gaussian-shaped pips of a frequency and duration similar to those used here [2140 Hz, 3-ms actual duration; (Hoke, 1974)]. Brief Gaussian-shaped pips are therefore likely to produce the initial coordinated burst of primary afferent activity that is believed to underlie the N_1 (e.g., Kiang *et al.*, 1965; Pfeiffer and Kim, 1972). Now, if each neuron has just enough time to fire a single spike in response to the Gaussian pip, as is already suspected to occur for clicks, then the sizes of the respective neuronal populations are represented by the total numbers of resulting spike counts. Generally, the spike count is unmeasurable, but count may be assumed to be proportional to the N_1 potential. If N_1 variance is assumed to not change for small level changes, a detectability index (*d*-prime) can be computed and used to estimate the DL (Nizami, 2000). When the means and variances of the click-evoked N_1 (Frishkopf, 1956) are put into the model, a W-shaped pattern of discriminability appears as a function of click level, with a profound mid-level rise at about 50 dB below maximum level. The DLs are otherwise of the order of magnitude known for man.

It cannot be assumed *a priori* that the mechanisms that support intensity discrimination at longer durations are completely absent at the short durations for which the mid-level hump occurs. The ANOVA for the Gaussian-shaped stimuli suggest that the observed mid-level hump for tones or noise is superimposed on some classical pattern of level-dependence of the DL. A clue to the origin of this behavior comes from the one study that showed peri-stimulus-time histograms of primary afferent neurons in response to brief tone-pips. Geisler and Sinex (1982) exposed cats to pure-tone bursts whose envelopes were isosceles triangles. An 8-ms tone-pip of 3.5 kHz evokes an initial burst of high probability, followed by a spike train. Geisler and Sinex counted the spikes evoked in each neuron, inherently assuming that firing-rate is the significant variable. Counts were made for tone-pips of fixed level but different frequency, then plotted versus frequency to obtain a ''response area'' for each level. The characteristic frequency (CF) of the neuron showed clearly as a peak in each such plot. These response plots were provided for three afferents, whose CFs were 0.7,

4.9, and 21 kHz. For CF=0.7 kHz, the peaks progressively lost their sharpness as duration dropped from 32 to 8 to 4 ms (i.e., there is spectral splatter). Shortening the tone to 2 ms removed any apparent frequency selectivity. In contrast, a burst of only 1 ms showed good frequency selectivity for the high-frequency afferent (CF=21 kHz). For the mid-frequency afferent, the results were intermediate, with good resolution for the 2-ms tone but poor resolution for the 1-ms burst.

It has been assumed for many years that firing-rates, derived by counting the peri-stimulus spikes evoked in individual primary afferents over the suspected integration times of the ear, are responsible for the coding of stimulus level and changes in stimulus level. Based on the findings of Geisler and Sinex (1982), the peri-stimulus spike train evoked in a single afferent by a Gaussian-shaped stimulus can be conceptually divided into two parts. Reliance on the coordinated initial burst produces the mid-level hump, as found for clicks, which have no peri-stimulus-time spike train. If the stimulus is longer than a click, for example, a Gaussian-shaped pip, then once the coordinated initial burst is past, there will be a consequent peri-stimulus-time spike train that provides a firing-rate code in individual neurons. But for brief pips, the mid-level hump predominates, presumably because the initial burst dominates the response. Firing-rate coding does not predominate until the ramping of the stimulus is sufficiently gradual to eliminate the coordination of the initial neuronal spike burst, or until the spike train reaches some critical duration, or both. Unfortunately the literature does not supply enough data to evaluate these hypotheses.

D. A central counting point?

A code for level based on total counts of spikes demands a central counting site, and there is some evidence for this concept from forward-masking experiments. Zeng *et al.* (1991) obtained intensity DLs for a 25-ms, 1-kHz tone presented 100 ms after a 90 dB SPL noise that was 200 Hz wide and centered on 1 kHz. The resulting mean DLs (see also Zeng and Turner, 1992; Plack *et al.*, 1995; Zeng and Shannon, 1995) are similar to those seen in Fig. 3 for the 2.51-ms 2-kHz tone-pip. The mid-level hump has been found under backward-masking, for which it is generally accepted that the masker cannot affect the neuronal response to the tone except within the limits of ringing. But Plack *et al.* (1995) found that the observed effect extends far beyond those limits, suggesting some kind of central interaction. Plack *et al.* (1995) also found a relatively small forward-masking effect when masker and tone were presented respectively to the different ears (see also Zeng and Shannon, 1995, and Schlauch *et al.*, 1999) as well as a contralateral backward-masking effect. These phenomena require a central interaction, because the first interaction of left- and right-ear inputs occurs in the brainstem (Irvine, 1986).

At first, explanations of the nonmasked mid-level rise of Figs. 3 and 5 in terms of peripheral neural behavior would appear to have nothing in common with possible explanations of the ''Zeng effect''. However, both types of mid-level humps may involve a process in which evoked spikes are summed at some central point so that encoding of intensity and intensity change can be based on spike counts from more than one neuron. This spike pooling is likely to occur; the discriminability offered by single primary afferent neurons is not good enough to account for psychophysical performance (see, for example, Nizami and Schneider, 1997; Nizami, 1998), and the existence of critical bands implies convergence of neuronal output at some higher counting locus or loci. One of these loci must be where left-ear input reduces the response of the collector neuron to right-ear input, and vice versa, one possible mechanism for contralateral masking.

The across-subject averages for the 2-kHz Gaussian-shaped tone-pip (Fig. 3) are not incompatible with the DLs of Carlyon and Moore (1984) for 6.5- and 8-kHz tones. This superficial resemblance should not be taken as an indicator of a common *peripheral* mechanism. It has been proposed here that the mid-level hump evoked by clicks or by Gaussian-shaped tone-pips is determined by the coordinated neural burst represented by the N_1 potential. In their experiments, Carlyon and Moore employed stimuli having a plateau duration of 26 ms, that is presumably large enough for a rate-based intensity code. Further, Carlyon and Moore used raised-cosine ramps of 5 ms, a duration long enough to eliminate the N_1 potential (Goldstein and Kiang, 1958; cat) and, by inference, any cues that might arise from the coordinated across-neurons burst represented by the existence of N_1. Caution is advised because Goldstein and Kiang employed broadband noise, not pure tones; for pure tones, ramps of fixed duration are effectively more gradual at higher frequencies.

The mid-level hump seen in long-tone DLs at high frequencies is not likely to arise from the same peripheral response that is postulated to account for the mid-level hump seen for much shorter stimuli. The phenomenon joining the two kinds of mid-level humps, if any, must therefore be central. The common factor may be the counting center that operates on the input from the periphery to create the forward-masked mid-level hump and the brief-stimulus mid-level hump.

V. SUMMARY AND CONCLUSIONS

Intensity-difference limens have been established as a function of level and duration for Gaussian-shaped 2-kHz tone-pips and for Gaussian-shaped packets of broadband noise. The pedestal levels spanned 30–90 dB SPL in 10-dB steps. The DLs for the 1.25- and 2.51-ms tone-pips show a mid-level rise at about 40 dB SL, whereas the DLs for tone-pips of 10.03 ms follow the near-miss to Weber's law, albeit with larger values than seen for longer durations (e.g., Jesteadt *et al.*, 1977). Analysis of variance shows a linear component, that presumably represents the near-miss, underlying all the DLs for the 2-kHz tone-pip. The DLs for the tone-pip are higher for 2.51-ms than for 1.25-ms stimuli, creating a mid-duration hump at 2.51 ms, consistent with recent reports (Baer *et al.*, 1999; van Schijndel *et al.*, 1999). Thus the data for tones could be regarded as revealing a mid-level hump at some critical range of durations, or a mid-duration hump at some critical range of levels.

Van Schijndel *et al.* (1999) have presented a model of the mid-duration hump for Gaussian-shaped tones, which was couched in terms of an interaction of bandwidth and duration effects at low sensation levels. The present results for the Gaussian-shaped noise-pip are consistent with the model, but suggest that it is better to think in terms of level and duration as the primary variables. The DLs for broadband-noise pips of $D=0.63$, 1.25, 2.51, 5.02, and 10.03 ms all show a significant mid-level hump at 30–40 dB SL. The level-dependence of the DL for the 0.63-ms noise-pips is comparable to that seen for clicks (Raab and Taub, 1969) despite the difference in their statistical properties. Analysis of variance reveals no linear component to the level-dependence, suggesting that Weber's law itself underlies all the noise-pip DLs. The DLs drop monotonically with increase in duration, but the DLs for the 10.03-ms pip are still higher than those for longer stimuli (e.g., Miller, 1947).

It is hypothesized that the existence of both the mid-level hump and of classical patterns of level-dependence for the Gaussian-shaped pip is a reflection of two mechanisms of level-coding, each using a different portion of the spike train that is evoked in the primary afferent neuron by the pip. Triangular-enveloped tone-pips evoke an initial spike burst of high probability, followed by a spike train that lasts as long as the stimulus (Geisler and Sinex, 1982). The spike train is hypothesized to produce a firing-rate code that causes the classical patterns of level-dependence of the DL. The neuron's initial spike burst, on the other hand, has the special property of being coordinated in time with the initial burst appearing in any other neuron. No method has been devised that allows this instantaneous massed spike count to be accurately measured, but it does produce a measurable epiphenomenon, the N_1 potential. It has been shown that the mid-level hump known for clicks can be predicted from the behavior of the N_1 potential (Raab and Taub, 1969; Taub and Raab, 1969), thus providing a working hypothesis of the origin of the mid-level hump.

ACKNOWLEDGMENTS

The measurements from Nizami (1999) were made at Erindale College, University of Toronto, and were supported in part by an NSERC grant to Bruce Schneider. The rest of the measurements were made at Boys Town National Research Hospital and were supported by a grant to Walt Jesteadt (R01 DC00136). We thank the reviewers, Chris Plack and Tammo Houtgast, for their many insightful comments.

[1]The Gaussian envelope was always smoothly reduced to zero amplitude at ± 4 standard deviations (σ), so that stimulus duration, as reported henceforth, was 8σ. The Gaussian-shaped tone-pips of Baer *et al.* (1999) were also cut-off at $\pm 4\sigma$, so the durations of the two pips can be directly compared. Baer *et al.* expressed duration as D where $D=\sigma\sqrt{2\pi}$, so that $8\sigma = 8D/\sqrt{2\pi} \approx 3.19D$. Multiplying D by the peak height of the Gaussian curve yields a value of unity. As unity is the area under any probability density function such as the Gaussian, D therefore specifies the rectangular envelope that encloses the same area as a Gaussian envelope. A different conversion is required for the durations specified by van Schijndel *et al.* (1999), who employed the unitless "shape factor" α. For a sinusoidal carrier of frequency f_0, $\alpha=(\sigma f_0\sqrt{2\pi})^{-1}$ and hence $D=(\alpha f_0)^{-1}$. The durations of the stimuli used in the present experiments, in milliseconds, are approximately $3.19\cdot(2\alpha)^{-1}$. However, the Gaussian envelope used by van Schijndel *et al.* was cut-off when its amplitude fell to 60 dB below peak, which corresponds to approximately $\pm 3.72\sigma$. Thus one of the present Gaussian-enveloped tone-pips would be roughly 7.6% longer than a tone-pip of van Schijndel *et al.* that has the same carrier and the same shape factor. Note that a shape factor can be specified only for a pure tone, so that Gaussian-enveloped broadband noise does not have a shape factor.

Abramovich S. J. (**1978**). "Hearing investigation in relation to the duration of acoustic signals," Laryngoscope, **88**, 334–341.
Avakyan, R. V., and Radionova, E. A. (**1963**). "The special features of differential intensity thresholds for a brief sound signal," Sov. Phys. Acoust. **8**, 320–323.
Baer, T., Moore, B. C. J., and Glasberg, B. R. (**1999**). "Detection and intensity discrimination of Gaussian-shaped tone pulses as a function of duration," J. Acoust. Soc. Am. **106**, 1907–1916.
Buus, S. (**1990**). "Level discrimination of frozen and random noise," J. Acoust. Soc. Am. **87**, 2643–2654.
Campbell, R. A. (**1963**). "Detection of a noise signal of varying duration," J. Acoust. Soc. Am. **35**, 1732–1737.
Carlyon, R. P., and Beveridge, H. A. (**1993**). "Effects of forward masking on intensity discrimination, frequency discrimination, and the detection of tones in noise," J. Acoust. Soc. Am. **93**, 2886–2895.
Carlyon, R. P., and Moore, B. C. J. (**1984**). "Intensity discrimination: a severe departure from Weber's law," J. Acoust. Soc. Am. **76**, 1369–1376.
De Boer, E. (**1966**). "Intensity discrimination of fluctuating signals," J. Acoust. Soc. Am. **40**, 552–560.
Florentine, M. (**1986**). "Level discrimination of tones as a function of duration," J. Acoust. Soc. Am. **79**, 792–798.
Florentine, M., and Buus, S. (**1981**). "An excitation-pattern model for intensity discrimination," J. Acoust. Soc. Am. **70**, 1646–1654.
Florentine, M., Buus, S., and Mason, C. R. (**1987**). "Level discrimination as a function of level for tones from 0.25 to 16 kHz," J. Acoust. Soc. Am. **81**, 1528–1541.
Frishkopf, L. S. (**1956**). "A probability approach to certain neuroelectric phenomena," M.I.T. Res. Lab. Electron. Tech. Rep. **307**, 1–74.
Gabor, D. (**1946**). "Theory of communication," J. Inst. Elec. Eng. **93**, 429–457.
Geisler, C. D., and Sinex, D. G. (**1982**). "Responses of primary auditory fibers to brief tone bursts," J. Acoust. Soc. Am. **72**, 781–794.
Goldstein, M. H., and Kiang, N. Y.-S. (**1958**). "Synchrony in neural activity in electric responses evoked by transient acoustic stimuli," J. Acoust. Soc. Am. **30**, 107–114.
Green, D. M. (**1960**). "Auditory detection of a noise signal," J. Acoust. Soc. Am. **32**, 121–131.
Green, D. M., and McGill, W. J. (**1970**). "On the equivalence of detection probabilities and well-known statistical quantities," Psychol. Rev. **77**, 294–301.
Hoke, M. (**1974**). "Influence of certain stimulus parameters on the compound action potential, as demonstrated at normal subjects and in some pathological cases," Rev. Laryngol. Otol. Rhinol. (Bord) **95**, 508–514.
Irvine, D. R. F. (**1986**). "The auditory brainstem. Chapter 5. Superior olivary complex: anatomy and physiology," Prog. Sens. Physiol. **7**, 79–121.
Jesteadt, W., Wier, C. C., and Green, D. M. (**1977**). "Intensity discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 169–177.
Keppel, G. (**1991**). *Design and Analysis* (Prentice–Hall, Englewood Cliffs, NJ).
Kiang, N. Y.-S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT, Cambridge, MA).
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
McGill, W. J., and Goldberg, J. P. (**1968**). "A study of the near-miss involving Weber's law and pure-tone intensity discrimination," Percept. Psychophys. **4**, 105–109.
Miller, G. A. (**1947**). "Sensitivity to changes in the intensity of white noise and its relation to masking and loudness," J. Acoust. Soc. Am. **19**, 609–619.
Moore, B. C. J., and Raab, D. H. (**1974**). "Pure-tone intensity discrimination: some experiments relating to the "near-miss" to Weber's law," J. Acoust. Soc. Am. **55**, 1049–1054.
Nizami, L. (**1998**). "Dynamic ranges of auditory afferents: little difference between sloping-saturating and sigmoidal units," Soc. Neurosci. Abs. **24**, 901.

Nizami, L. (**1999**). ''On Auditory Dynamic Range,'' Doctoral dissertation, University of Toronto.
Nizami, L. (**2000**). ''The level-dependence of the intensity-difference limens for very brief stimuli, inferred from the N1 potential in the cat,'' ARO Abs. **23**, 27.
Nizami, L., and Schneider, B. A. (**1994**). ''Forward-masked intensity increment thresholds at two recovery times,'' J. Acoust. Soc. Am. **96**, 3280.
Nizami, L., and Schneider, B. A. (**1997**). ''Auditory dynamic range derived from the mean rate-intensity function in the cat,'' Math. Biosci. **141**, 1–28.
Nizami, L., and Schneider, B. A. (**1999**). ''The fine structure of the recovering auditory detection threshold,'' J. Acoust. Soc. Am. **106**, 1187–1190.
Pfeiffer, R. R., and Kim, D. O. (**1972**). ''Response patterns of single cochlear nerve fibers to click stimuli: descriptions for cat,'' J. Acoust. Soc. Am. **52**, 1669–1677.
Plack, C. J. (**1998**). ''Beneficial effects of notched noise on intensity discrimination in the region of the 'severe departure','' J. Acoust. Soc. Am. **103**, 2530–2538.
Plack, C. J., Carlyon, R. P., and Viemeister, N. F. (**1995**). ''Intensity discrimination under forward and backward masking: Role of referential coding,'' J. Acoust. Soc. Am. **97**, 1141–1149.
Raab, D. H., and Taub, H. B. (**1969**). ''Click-intensity discrimination with and without a background masking noise,'' J. Acoust. Soc. Am. **46**, 965–968.
Schlauch, R. S., Clement, B. R., Ries, D. T., and DiGiovanni, J. J. (**1999**). ''Masker laterality and cueing in forward-masked intensity discrimination,'' J. Acoust. Soc. Am. **105**, 822–828.
Taub, H. B., and Raab, D. H. (**1969**). ''Fluctuations of N_1 amplitude in relation to click-intensity discrimination,'' J. Acoust. Soc. Am. **46**, 969–978.
van Schijndel, N. H., Houtgast, T., and Festen, J. M. (**1999**). ''Intensity discrimination of Gaussian-windowed tones: indications for the shape of the auditory frequency-time window,'' J. Acoust. Soc. Am. **105**, 3425–3435.
Viemeister, N. F. (**1988**). ''Psychophysical aspects of auditory intensity coding,'' in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, Toronto), pp. 213–241.
Zeng, F.-G., and Shannon, R. V. (**1995**). ''Possible origins of the nonmonotonic intensity discrimination function in forward masking,'' Hear. Res. **82**, 216–224.
Zeng, F.-G., and Turner, C. W. (**1992**). ''Intensity discrimination in forward masking,'' J. Acoust. Soc. Am. **92**, 782–787.
Zeng, F.-G., Turner, C. W., and Relkin, E. M. (**1991**). ''Recovery from prior stimulation II: effects upon intensity discrimination,'' Hear. Res. **55**, 223–230.
Zwicker, E. (**1970**). ''Masking and psychological excitation as consequences of the ear's frequency analysis,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), pp. 376–396.

The variation across time of sensitivity to interaural disparities: Behavioral measurements and quantitative analyses

Michael A. Akeroyd[a)]
Department of Neuroscience, University of Connecticut Health Center, Farmington, Connecticut 06030 and Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, United Kingdom

Leslie R. Bernstein
Department of Neuroscience, University of Connecticut Health Center, Farmington, Connecticut 06030

(Received 11 December 2000; revised 3 May 2001; accepted 15 August 2001)

Zurek (1980) measured listeners' sensitivities to interaural disparities conveyed by a 5-ms ''probe'' segment embedded within a 50-ms burst of otherwise diotic broadband noise [P. M. Zurek, J. Acoust. Soc. Am. **67**, 952–964 (1980)]. He found that thresholds for interaural time delay (ITD) and interaural intensitive difference (IID) were markedly elevated when the onset of the probe segment occurred between 1 and 5 ms after the onset of the burst. Zurek postulated that this occurred because the leading portion of the noise briefly inhibited sensitivity to subsequent binaural information. If such inhibition were the primary factor responsible for the elevation in thresholds, then the omission of the portion of the noise trailing the probe segment would be expected to have little, if any, influence on performance. In order to test this hypothesis, listeners' sensitivities to ITD and IID were measured using a paradigm similar to that employed by Zurek. The results revealed that the omission of either the leading or the trailing portions of the diotic noise led to substantial reductions in threshold ITDs and IIDs. The data were successfully accounted for by a model based upon a combination of a temporal window with an equivalent rectangular duration of approximately 10 ms and a weighting function representing a brief loss of binaural sensitivity just after the onset of a sound. [DOI: 10.1121/1.1412442]

PACS numbers: 43.66.Pn, 43.66.Mk, 43.66.Ba [SPB]

I. INTRODUCTION

A number of investigations have demonstrated that the binaural system appears to average, or smooth over, dynamic changes in interaural disparities. In general, the behavioral data have been accounted for quantitatively by assuming that interaural information is passed through a ''temporal window,'' the form of which specifies the averaging of the binaural information (e.g., Grantham and Wightman, 1978; Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998; Akeroyd and Summerfield, 1999; Bernstein *et al.*, 2001). While it is the case that binaural information appears to be averaged over time, it seems that not all portions of the stimuli are weighted equally. Rather, it appears that sensitivity to binaural information is diminished approximately 5–10 ms after the onset of a sound (Zurek, 1980, 1987; Freyman *et al.*, 1991; Aoki and Houtgast, 1992; Houtgast and Aoki, 1994; Tollin and Henning, 1998). Zurek (1987) combined these notions in a conceptual model to account, qualitatively, for data obtained using a variety of experimental paradigms designed to investigate the precedence effect and related phenomena (e.g., Blauert, 1997; Litovsky *et al.*, 1999). Specifically, the model incorporated both a brief ''post-onset'' loss of binaural sensitivity followed by a mechanism to average binaural information over time.

The present study was motivated by Zurek's (1980) measurements of listeners' sensitivity to interaural disparities conveyed by a 5-ms burst of broadband noise (the ''probe segment'') embedded within a 50-ms burst of otherwise diotic noise. He observed that thresholds for interaural temporal difference (ITD) and interaural intensitive difference (IID) were markedly elevated when the onset of the probe segment occurred between 1 and 10 ms after the onset of the diotic noise. Zurek accounted for his results by postulating that the leading portion of the diotic noise briefly inhibited sensitivity to the binaural information conveyed by the probe segment.

If inhibition, or a post-onset loss of sensitivity, arising from the leading portion of the noise was the primary factor determining the elevated thresholds that Zurek (1980) observed, then the trailing portion of the diotic noise (that is, the portion following the probe segment) would be expected to have little or no influence on thresholds. In order to test this hypothesis, we used Zurek's paradigm and included additional conditions in which we omitted either the leading or the trailing portions of the diotic noise. Our results suggest that the elevation in thresholds that Zurek observed was not determined solely by a brief loss of sensitivity after the onset of the stimulus. Rather, as will be shown, the best description of our data resulted from applying a model that, like Zurek's (1987) conceptual model, incorporated both a temporal window and a weighting function that described a brief ''post-onset'' loss of binaural sensitivity.

II. EXPERIMENT

A. Design

The listeners' task was to detect an interaural disparity (either ITD or IID) conveyed by a 5-ms probe segment of an otherwise diotic burst of broadband noise. Figure 1 depicts

a)Electronic mail: maa@biols.susx.ac.uk

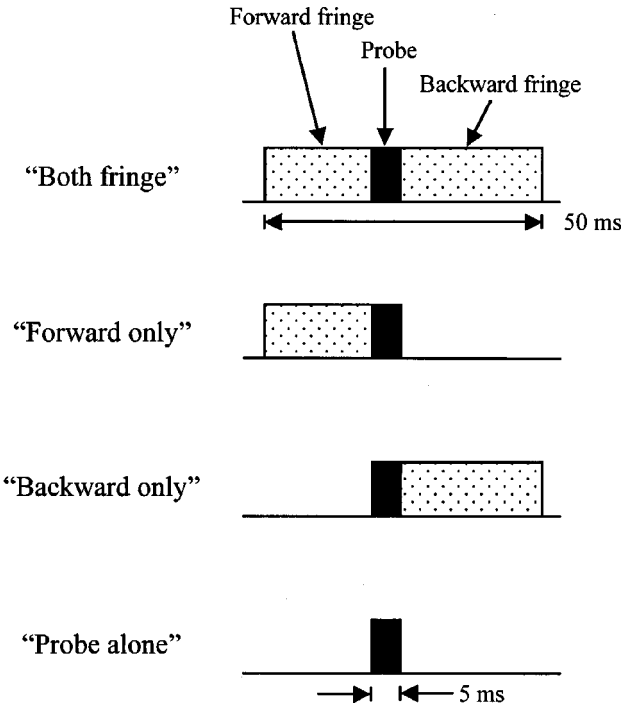


FIG. 1. Schematic illustration of the four sets of stimulus conditions used in the experiment. In the ''both-fringe'' conditions (first row), the probe segment is embedded within two diotic bursts. The ''forward-only'' conditions (second row) differ from the ''both-fringe'' conditions in that the diotic burst occurring *after* the probe is removed. The ''backward-only'' conditions (third row) differ from the ''both-fringe'' conditions in that the diotic burst occurring *before* the probe is removed. In the ''probe-alone'' condition (fourth row), the probe is presented in isolation.

the four sets of conditions employed. The ''both-fringe'' conditions were essentially a replication of those used by Zurek (1980). In these conditions, the total duration of the noise was 50 ms and the temporal position of the probe segment was varied. The ''forward-only'' conditions were the same as the both-fringe conditions except that the *trailing* portion of the diotic noise was omitted; the ''backward-only'' conditions were the same as the both-fringe conditions except that the *leading* portion of the diotic noise was omitted. In the ''probe alone'' condition, the stimulus consisted of the 5-ms probe segment of noise presented in isolation. Table I lists the values of the temporal parameters employed for each condition. Consistent with the approach employed by Zurek (1980), the functions relating threshold to the onset delay of the probe (i.e., the duration of the forward fringe) were measured with a finer resolution for onset-delays less than 10 ms as compared to that used for onset-delays greater than 10 ms.

B. Methods

1. Stimuli and apparatus

Each burst of noise presented to the listeners was formed by concatenating independent tokens of noise for the forward-fringe, probe, and backward-fringe segments. The segments were constructed in the frequency domain with a sampling rate of 20 kHz (TDT AP2). The magnitudes of components below 8 kHz had a Rayleigh distribution and the magnitudes of all other components were set to zero. The phases of all components were sampled randomly from a uniform distribution. ITDs were applied to the probe segment by appropriate linear phase shifts of the components; IIDs were applied by symmetric scaling of the magnitudes of the components (i.e., an increase in one ear, a decrease in the other). An inverse FFT was applied to the resulting spectra to yield 8192-point waveforms in the time domain, from which segments of the desired durations were extracted. For the (5-ms) probe segments, new, independent waveforms were generated for each interval of each trial. For the (generally longer) forward-fringe and backward-fringe segments, however, five independent waveforms were generated per trial. For each of the four intervals composing a trial, two of these five noises were chosen at random and utilized for the forward and backward fringes that temporally surrounded the probe. The transitions between the segments that were concatenated to form each burst of noise were essentially instantaneous. Cos^2 ramps having an overall duration of 1 ms were then applied to the onset and offset of each burst of noise. The bursts were converted to electrical waveforms by a TDT PowerDac with a resolution of 16 bits. All waveforms were low-pass filtered at 8.5 kHz (TDT FLT2), passed through programmable attenuators (TDT PA2) and presented via matched TDH-39 earphones mounted in MX/41-AR cush-

TABLE I. The values of the temporal parameters (in ms) defining each of the conditions tested in the experiment. In all conditions the duration of the probe segment was 5 ms. A ''probe-alone'' condition was also tested, for which the stimulus was the same 5-ms probe segment but presented in isolation.

Both-fringe conditions				Forward-fringe conditions			Backward-fringe conditions		
Onset delay of probe	Duration of forward fringe	Duration of backward fringe	Overall duration	Onset delay of probe	Duration of forward fringe	Overall duration	Onset delay of probe	Duration of backward fringe	Overall duration
0.0	0.0	45.0	50.0						
0.5	0.5	44.5	50.0	0.5	0.5	5.5	0.0	44.5	49.5
1.0	1.0	44.0	50.0	1.0	1.0	6.0	0.0	44.0	49.0
2.5	2.5	42.5	50.0	2.5	2.5	7.5	0.0	42.5	47.5
5.0	5.0	40.0	50.0	5.0	5.0	10.0	0.0	40.0	45.0
7.5	7.5	37.5	50.0	7.5	7.5	12.5	0.0	37.5	42.5
10.0	10.0	35.0	50.0	10.0	10.0	15.0	0.0	35.0	40.0
15.0	15.0	30.0	50.0	15.0	15.0	20.0	0.0	30.0	35.0
25.0	25.0	20.0	50.0	25.0	25.0	30.0	0.0	20.0	25.0
35.0	35.0	10.0	50.0	35.0	35.0	40.0	0.0	10.0	15.0

ions to listeners seated in single-walled IAC booths. All the stimuli were presented at a nominal spectrum level of 45 dB SPL (i.e., before applying any IID).

2. Procedure

The procedure was identical to that used by Bernstein *et al.* (2001). A four-interval, two-cue, two-alternative temporal forced-choice adaptive task was used. Each trial consisted of a warning interval (500 ms) and four observation intervals separated by 450 ms. The observation intervals were marked by a visual display on a computer monitor. In the first and fourth ''cueing'' intervals, the forward-fringe, probe, and backward-fringe segments were all diotic. The interaural disparity to be detected (either ITD or IID) was imposed on the probe segment in either the second or the third interval with equal *a priori* probability. The interaural conditions in the remaining interval were identical to those in the cueing intervals. Feedback was provided for approximately 400 ms after the listener responded. When the listeners' task was to detect an IID, each nonsignal interval contained either an increment at both ears or a decrement at both ears (with equal *a priori* probability) that were equal in magnitude to the symmetric increments/decrements that composed the IID in the signal interval. These increments and decrements occurred at identical times and for identical durations as the symmetric increments/decrements imposed during the signal interval. This procedure was followed in order to prevent listeners from solving the IID-detection task on the basis of monaural, energy-related cues.

The ITD or IID of the probe segment was varied adaptively in order to estimate the 70.7%-correct point on the psychometric function (Levitt, 1971). When threshold ITDs were measured, the initial step size for the adaptive track corresponded to a factor of 1.585 (a ratio equivalent to 0.2 log units) and was reduced to a factor of 1.122 (a ratio equivalent to 0.05 log units) after two reversals (see Saberi, 1995). When threshold IIDs were measured, the initial and final step sizes were 1.0 and 0.5 dB, respectively. For both the measurement of threshold ITDs and threshold IIDs, the adaptive track was terminated after 12 reversals. Threshold ITDs were defined as the geometric mean of the value of the ITD (in μs) at each of the last ten reversals. Threshold IIDs were defined as the arithmetic mean of the value of the IID (in dB) at each of the last ten reversals. For each listener and condition, final estimates of thresholds were calculated by averaging the individual thresholds obtained from six adaptive runs.

Testing began with measurement of threshold ITDs. For each listener, three thresholds were obtained for each of the conditions listed in Table I, chosen in random order. Then three more thresholds were obtained at each condition, with the conditions visited in the reverse order. This entire process was repeated for the measurement of threshold IIDs. For one listener, threshold IIDs were measured before threshold ITDs. All of the listeners were given at least 5 h of practice before formal collection of data began.

3. Listeners

Five listeners, three females and two males, participated in the experiment. Threshold ITDs were obtained for all five listeners. Only three of the listeners were available for the measurement of threshold IIDs. All had served in previous experiments. None had thresholds higher than 15 dB HL at octave frequencies between 250 Hz and 8000 Hz, inclusive (ANSI, 1969). The listeners were paid for their participation.

C. Results and discussion

The top half of Fig. 2 (panels A and B) displays the threshold ITDs. The bottom half (panels C and D) displays the threshold IIDs. The left-hand panels display the thresholds for the both-fringe (squares) and forward-only (asterisks) conditions and are plotted as a function of the duration of the forward fringe (see Table I). The right-hand panels display the thresholds for the both-fringe (squares) and backward-only (diamonds) conditions. Note that the data in the right-hand panels are plotted as a function of the duration of the backward fringe (see Table I), with the result that the order of the both-fringe data is reversed relative to that in the left-hand panels. In each panel, the dashed line represents the thresholds in the probe-alone condition. Because the data obtained from each listener exhibited similar trends, each panel displays the average thresholds across the listeners. For clarity, individual error bars are not displayed. Rather, two error bars are drawn at the top-right corner of each panel. They were computed by averaging the individual standard errors of each of the plotted thresholds for a particular condition. For example, in panel A, the standard error, across listeners, was computed for each of the eleven both-fringe conditions (squares). Those values were then averaged. The error bars extending from the square in the upper right-hand corner of panel A span twice that averaged value.

As will be discussed in detail below, the data depicted in Fig. 2 reveal three primary results. First, the both-fringe data are consistent with those obtained by Zurek (1980). Second, the removal of either the leading or the trailing diotic fringes leads to substantial reductions in thresholds. Third, compared to the probe-alone condition, threshold ITDs are elevated to a greater extent by the presence of a short forward fringe than by a short backward fringe.

Beginning with the both-fringe conditions in panels A and C (squares), the data obtained from our listeners revealed effects consistent with those obtained by Zurek (1980, Fig. 8). In general, as the duration of the forward fringe was increased from 0 ms, thresholds increased rapidly, reaching a maximum at a forward-fringe duration of about 5–10 ms. Then, as the duration of the forward fringe was increased further, thresholds declined gradually.

In all cases, threshold ITDs and threshold IIDs obtained in the forward-only-conditions (asterisks) were markedly smaller than their counterparts obtained in the both-fringe conditions. That is, removing the portion of diotic noise that occurred *after* the probe segment resulted in decreases in both threshold ITDs and threshold IIDs. This new finding is inconsistent with the hypothesis that the elevated thresholds observed in the both-fringe conditions result primarily from

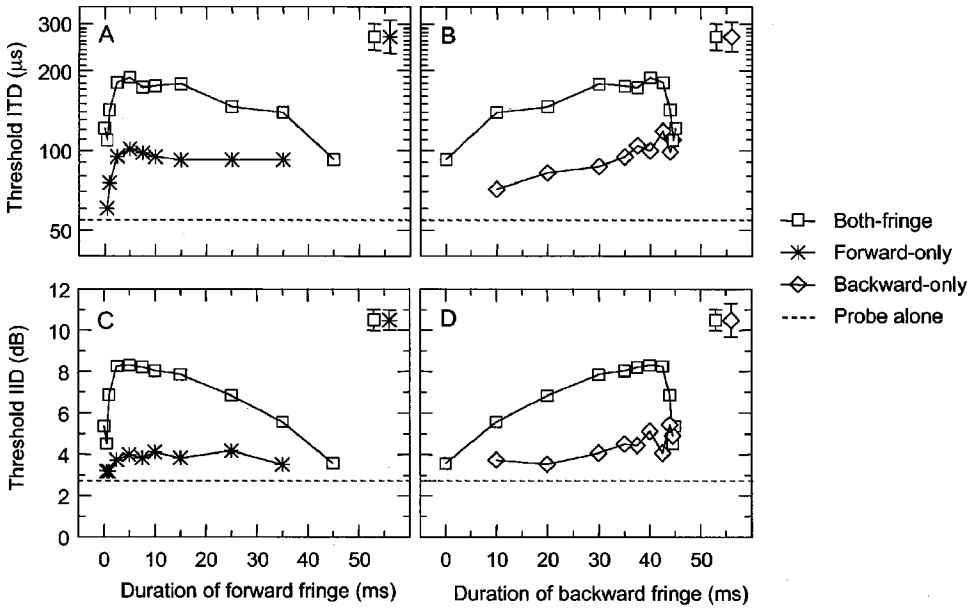


FIG. 2. **Top row**: Threshold ITD (in μs) as a function of the duration of the forward fringe (panel A) and as a function of the duration of the backward fringe (panel B). **Bottom row**: Threshold IID (in dB) as a function of the duration of the forward fringe (panel C) and as a function of the duration of the backward fringe (panel D). The parameter within each plot is the experimental configuration. Both-fringe thresholds are plotted as squares; forward-only thresholds are plotted as asterisks; backward-only thresholds are plotted as diamonds. The data points represent the average thresholds across five listeners (threshold ITDs) or three listeners (threshold IIDs). The two error bars plotted at the top-right corner of each panel span twice the average of the standard errors for each of the two conditions plotted in the panel. The average standard error was computed by calculating the mean of the individual standard errors of each of the plotted thresholds for a particular condition.

some type of inhibition produced by the leading portion of the diotic noise.

Panels B and D show the results for the both-fringe (squares) and backward-only (triangles) conditions. The threshold ITDs and threshold IIDs obtained in the backward-only conditions were, like those obtained in the forward-only conditions, markedly smaller than their counterparts obtained in the both-fringe conditions. That is, removing the portion of diotic noise that occurred *before* the probe segment, like removing the portion after the probe segment (see Fig. 1), resulted in a decrease in both threshold ITDs and threshold IIDs. Thus neither the presence of the diotic noise before, nor the presence of the diotic noise after, the probe segment is sufficient to account for the magnitude of the thresholds obtained in the both-fringe conditions. It is important to note, however, that the presence of diotic noise before the probe in the forward-only condition still produces a relatively steep increase in both threshold ITDs and IIDs as a function of the duration of the fringe. The difference between the rate of increase for the forward-only and the backward-only conditions is especially marked for threshold ITDs.

III. QUANTITATIVE ACCOUNTS OF THE DATA

A. Temporal window analysis

In these first computational analyses, we consider the degree to which the use of temporal windows could account for the threshold ITDs. In order to illustrate how the operation of a temporal window would be expected to affect thresholds, we describe three cases. First, consider one of the both-fringe conditions in which the probe segment is preceded and followed by diotic noise. Assuming the operation of a temporal window that overlaps the leading and trailing portions of the diotic noise, the values of ITD within the probe segment will be integrated, or averaged, with the zero values of ITD within the diotic noise. The result is that the binaural information conveyed by the probe segment will be effectively ''diluted'' by the binaural information conveyed by the diotic noise. Thus the average ITD measured at the output of the window will be less than the ITD of the probe segment alone. One consequence of this averaging is that, in order to reach threshold, a *larger* ITD would have to be applied to the probe segment than would be the case if the probe segment were presented in isolation. These expectations are consistent with the data shown in panel A of Fig. 2 (squares vs dashed line).

Second, consider another of the both-fringe conditions, specifically, one in which the temporal position of the probe segment is close to the onset (or offset) of the stimulus. Under the assumption that the temporal window is essentially centered on the probe segment, the total amount of diotic noise integrated within the window will be greatly reduced. This would lead to a lower threshold ITD than that obtained for the case described above. This account is consistent with the finding that thresholds in the both-fringe condition were lowest when the probe segment occurred close to the onset or offset of the stimulus (panel A of Fig. 2, squares).

Third, consider the forward-only or backward-only conditions in which the trailing or leading portions, respectively, of the diotic noise were omitted. In both cases, because a portion of the diotic noise is removed, the degree to which the averaging performed by the temporal window will dilute the ITD of the probe segment will be reduced. This would,

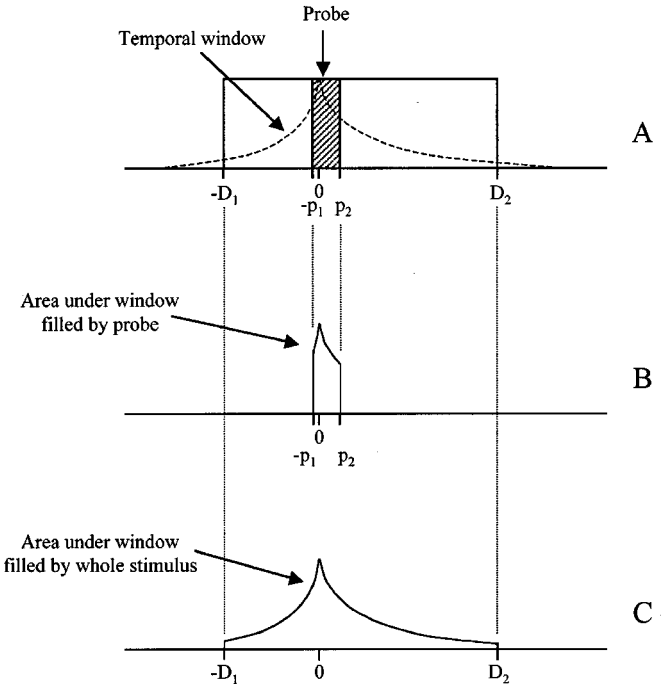


FIG. 3. Schematic illustration of the operation of the temporal window. Panel A depicts a both-fringe condition together with a temporal window (dashed line). The burst of noise begins at $-D_1$ and ends at D_2. The probe segment begins at $-p_1$ and ends at p_2. Panel B shows the area under the temporal window occupied by the probe segment. Panel C shows the area under the temporal window that is occupied by the whole stimulus. The output of the temporal window is assumed to be proportional to the ratio of these two areas. The times are measured relative to the peak of the window.

again, lead to lower threshold ITDs than expected for the first case described. This expectation is in accord with the data shown in panels A and B of Fig. 2 (squares vs asterisks; squares vs diamonds).

We now turn to a quantitative evaluation of the data. The general method was similar to that used by Bernstein *et al.* (2001). In order to fit the threshold ITDs, the operation of the temporal window was assumed to integrate the values of the ITD (in μs) conveyed by the probe segment with those conveyed by the surrounding noise. Because the surrounding noise was diotic (i.e., its ITD was zero), the output of the window, W, can be expressed by

$$W = \text{probe ITD} \frac{\int_{-p_1}^{p_2} w(\tau)\, d\tau}{\int_{-D_1}^{D_2} w(\tau)\, d\tau}, \tag{1}$$

where τ is time relative to the peak of the window and $w(\tau)$ is the function defining the shape of the window. The total duration of the probe segment is $|p_1|+|p_2|$ and the total duration of the stimulus is $|D_1|+|D_2|$. Panel A of Fig. 3 illustrates these temporal intervals. The numerator of Eq. (1) represents the area under the temporal window that is occupied by the probe segment (Fig. 3, panel B). The denominator represents that area that is occupied by the entire stimulus (Fig. 3, panel C). The effect of Eq. (1) is therefore to weight the ITD carried by the probe segment by the ratio of these two areas.

The form of the function defining $w(\tau)$, the shape of the temporal window, was

$$w(\tau) = \begin{cases} e^{\tau/T_1} & \tau < 0 \\ e^{-\tau/T_2} & \tau \geqslant 0, \end{cases} \tag{2}$$

where T_1 and T_2 are the time constants determining the response of the temporal window at times τ before and after, respectively, the occurrence of its peak response (at $\tau=0$). The temporal position of the window was allowed to vary in relation to the temporal position of the probe segment in order to maximize W. This was accomplished by calculating W as the peak of the window was moved between the onset and the offset of the probe segment, in 0.1-ms steps.

Next, it was assumed that the threshold ITDs measured in the experiment all corresponded to a constant, ''effective'' ITD, W_0, at the output of the window. Thus by re-arranging Eq. (1), the predicted threshold ITD is given by

$$\text{probe ITD at threshold} = W_0 \frac{\int_{-D_1}^{D_2} w(\tau)\, d\tau}{\int_{-p_1}^{p_2} w(\tau)\, d\tau}. \tag{3}$$

The MATLAB Optimization Toolbox function ''lsqnonlin'' was used to find the values of the time constants T_1 and T_2 of the window function and the value of W_0 (the effective ITD) that minimized the mean-squared error between the predictions and the experimental data in the both-fringe, forward-only, and backward-only conditions. The mean-squared errors were computed using the logs of the values of ITD as the dependent variable. All of the integrals were evaluated analytically.

The lines plotted in panel A of Fig. 4 represent the predicted thresholds for the both-fringe (solid line) and forward-only (dashed line) conditions. The symbols represent the behavioral thresholds and are replotted from panel A of Fig. 2. The equations that best fit these thresholds are shown to the right of panel A. The values of the time constants (4.1 and 5.1 ms) indicate that the best-fitting window is essentially symmetric. The predictions for the both-fringe conditions capture the general increase in threshold ITDs as the probe segment is moved away from the onset (or offset) of the stimulus. The predictions, however, substantially underestimate the thresholds ITDs for forward-fringe durations less than about 10 ms. As a result, the predictions fail to account for the steepness of the slope relating threshold ITDs to the duration of the fringe in this region. The predictions for the corresponding forward-only conditions conform well to the observed *overall* reduction in threshold ITDs (asterisks). Although not shown, the predictions for the backward-only conditions were also uniformly lower than the corresponding values obtained in the both-fringe conditions. Overall, the predictions accounted for 76% of the variance in the both, forward-only, and backward-only thresholds.[1]

The predictions derived from the use of the asymmetric temporal window did not capture the rapid rise in thresholds as the duration of the forward fringe was increased from 0 to about 10 ms. In order to determine whether this failure was peculiar to the choice of the asymmetric exponential window defined by Eq. (2), we repeated the analyses described above for threshold ITDs using an asymmetric rounded exponential (''roex'') and an asymmetric Gaussian window. These window functions were employed in the studies cited earlier

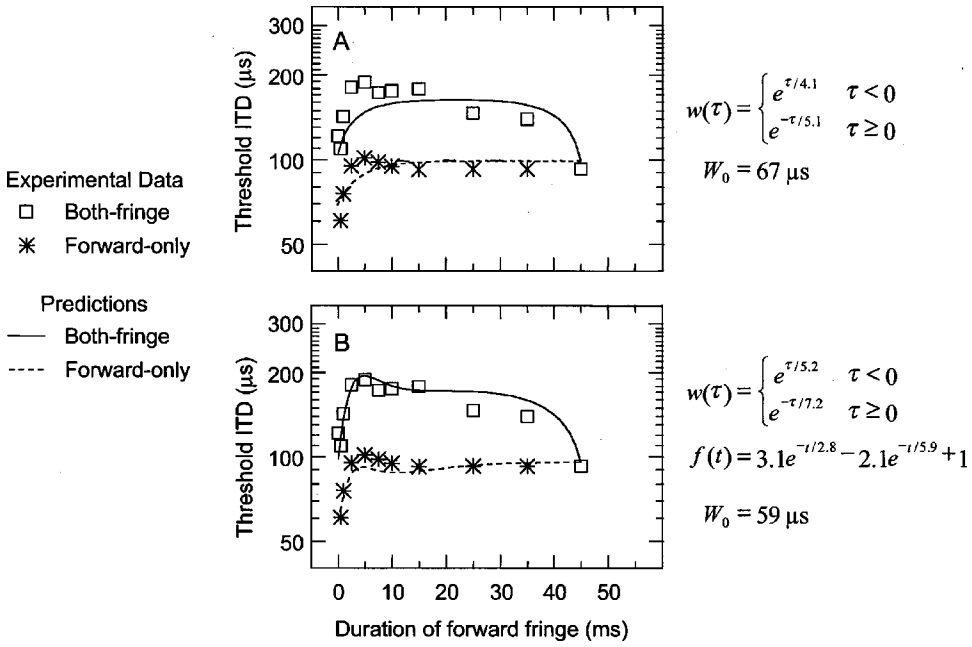


FIG. 4. ITDs (in μs) as a function of the duration of the forward fringe. The symbols represent the behavioral thresholds obtained in the both-fringe (squares) and forward-only (asterisks) conditions. The data are replotted from Fig. 2A. The lines represent predicted values for the both-fringe (solid line) and forward-only (dashed line) conditions. In panel A the predictions were derived using an asymmetric temporal window (Sec. III A). In panel B the predictions were derived using a post-onset-weighting function in combination with an asymmetric temporal window (Sec. III B). The equations resulting from the values of the parameters that best fit the thresholds are shown to the right of each panel. The unit of τ is milliseconds. The small tick marks on the *y*-axis are at 10-μs intervals.

concerning "binaural sluggishness" (e.g., Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998). The predictions obtained using these window functions, as well as the variance accounted for, were similar to those described above for the asymmetric exponential window. Indeed, none of the window functions tested was able to account for the rapid rise in thresholds as the duration of the forward fringe was increased from 0 to about 10 ms. That is, it appears that for these short durations of the forward fringe, thresholds deteriorate to a greater degree than is expected on the basis of the operation of a temporal window alone.

B. The addition of a post-onset weighting mechanism

In this second analysis, we re-visit the notion that some mechanism limits binaural sensitivity for a brief period after the onset of a stimulus. Specifically, we sought to test Zurek's (1987) conceptual model of the precedence effect and related phenomena in which a post-onset weighting function operates in tandem with a temporal window. In order to generate quantitative predictions of our data, we adopted Houtgast and Aoki's (1994, Fig. 5) equation describing the post-onset weighting, the general form of which is

$$f(t)=ae^{-t/T_a}+be^{-t/T_b}+1, \tag{4}$$

where t represents the time since the onset of the stimulus, a and b are constants and T_a and T_b are time constants. Houtgast and Aoki measured the relative salience of binaural cues within the leading and trailing portions of bursts of noise. The values of the parameters of $f(t)$ that they reported ($a=5$, $b=-2$, $T_a=2$, and $T_b=10$) were such that the weighting was greatest at the onset, rapidly decreased to a minimum after about 5 ms after the onset, and recovered to an intermediate value after about 20 ms.

In order to fit the threshold ITDs in Fig. 2, an asymmetric temporal window was assumed to integrate values of ITD resulting from the multiplication of the ITDs within the stimulus by the post-onset-weighting function. The form of the window was the same as that specified in Eq. (2). The MATLAB Optimization Toolbox function "lsqnonlin" was used to find the values of the parameters a, b, T_a and T_b of the post-onset weighting function, the values of the time constants T_1 and T_2 of the window function [see Eq. (2)], and the value of W_0 (the effective ITD) that minimized the mean-squared error between the predictions and the experimental data in the both, forward-only and backward-only conditions. All other aspects of the analysis were the same as those described earlier.

The lines plotted in panel B of Fig. 4 represent the predicted thresholds for the both-fringe (solid line) and forward-only (dashed line) conditions. In contrast to the predictions shown in panel A, the addition of the post-onset-weighting function resulted in the successful fitting of the steepness of the slope relating thresholds to the duration of the forward fringe for forward fringe durations less than about 10 ms. Furthermore, as can be seen, the inclusion of the post-onset-weighting mechanism did not compromise the other successful aspects of the fits that were obtained with only an asymmetric temporal window. The equations that best fit the thresholds are shown to the right of panel B, and the parameters are summarized in the first line of Table II. Once again, the values of the time constants indicate that the best-fitting window is essentially symmetric. Overall, the predictions accounted for 89% of the variance in the threshold ITDs in the

TABLE II. Values of the parameters of the temporal window for the two-stage model consisting of a post-onset-weighting mechanism followed by a temporal window (Sec. II B).

Task	T_1	T_2	W_0
ITD	5.2 ms	7.2 ms	59 μs
IID	4.1 ms	7.4 ms	2.7 dB

both-fringe, forward-only, and backward-only conditions.

As may be seen from Fig. 4, the values of the parameters of the temporal window differ only slightly across the two analyses. Furthermore, the values of the parameters of the post-onset weighing function differ little from those reported by Houtgast and Aoki (1994), attesting to the generality of their observations. In summary, it appears, as suggested by Zurek (1987), that the combination of a temporal window with a post-onset-weighting mechanism accounts well for the trends observed across our threshold ITDs.

Figure 5 displays the predictions derived from this model for all of the conditions included in the experiment. The layout of the panels correspond to that used in Fig. 2. Within each panel, the lines represent the predicted values; the symbols represent the behavioral thresholds. Panel A shows the predictions for the threshold ITDs that were described above and which were shown in panel B of Fig. 4. Panel B shows the predictions for the both-fringe (solid lines) and backward-only (dashed-line) conditions. (Recall that the values along the abscissa represent the duration of the backward fringe, with the result that the both-fringe data and the corresponding predictions are reversed relative to that in panel A.) As can be seen, the predictions of the backward-only conditions represent a successful account of the data.

Panels C and D of Fig. 5 depict the predictions for the threshold IIDs. Threshold IIDs were fitted in a similar manner to that used for threshold ITDs, but which differed in two respects. First, the temporal window was assumed to integrate the values of the IID (in dB) carried by the probe segment with the IIDs carried by the surrounding diotic noise. Second, the mean-squared errors were computed using the values of IID in dB as the dependent variable. The relations between the predictions and the data are essentially the same as those described earlier for threshold ITDs. In particular, the predictions conform well to the rapid rise in thresholds observed in the both-fringe conditions as the probe segment is moved away from the onset of the stimulus. Overall the predictions shown in panels C and D account for 92% of the variance in the threshold IIDs.

The values of the parameters of the temporal window resulting from the fitting procedure for the threshold IIDs are reported on the second line of Table II. The values of the time constants, T_1 and T_2, are remarkably similar to those reported above for threshold ITDs, which are listed in the first line of the table. This outcome is consistent with the findings of Bernstein *et al.* (2001), who concluded that a single temporal window could account for both threshold ITDs and threshold IIDs measured in a similar task.

The values of the parameters of the post-onset weighting function resulting from the fit to the threshold IID are reported in rows 1 and 2 of Table III. This function is depicted in Fig. 6 (dashed line) along with the corresponding function obtained for threshold ITDs (solid line). The two functions conform well to the general form described by Houtgast and Aoki (1994), in that the weighting was greatest at the onset, rapidly decreased to a minimum after about 5 ms after the onset, and recovered to an asymptotic value (a weight of 1.0) after about 20 ms. Furthermore, our data, like Houtgast and Aoki's, imply that the duration of the region of "suppressed weight" (i.e., weights less than 1.0) is slightly greater for the processing of IIDs than for the processing of ITDs.

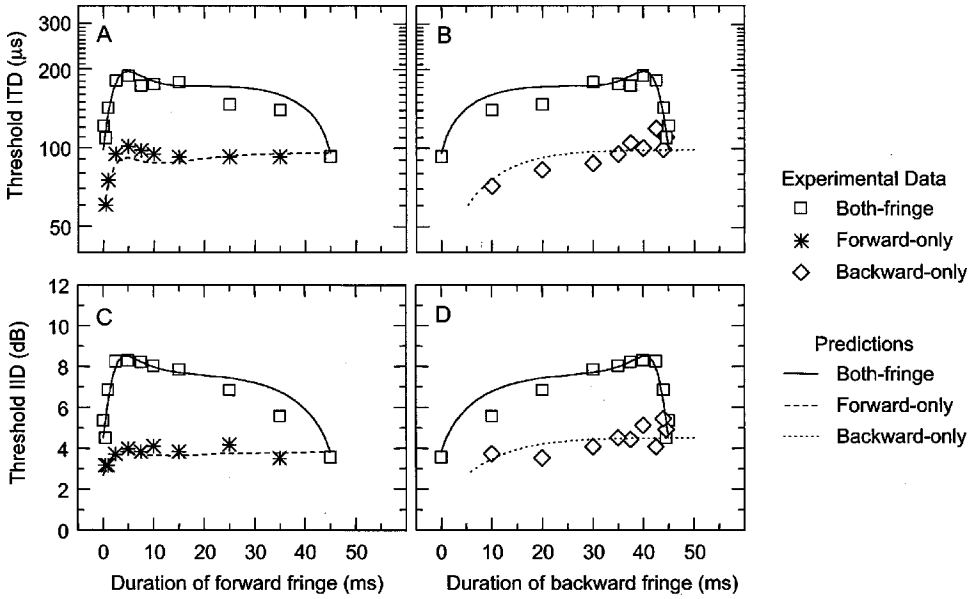


FIG. 5. Same as Fig. 2, with the exception that the lines represent the predictions derived using a post-onset-weighting function in combination with an asymmetric temporal window (Sec. III B). The solid line represents predictions for the both-fringe conditions, the dashed lines represent predictions of the forward-only conditions, and the dotted lines represent predictions for the backward-only conditions. Note that panel A of this figure is identical to panel B of Fig. 4.

TABLE III. Values of the parameters of the post-onset-weighting mechanism for the two-stage models consisting of (1) a post-onset-weighting mechanism followed by an asymmetric temporal window (upper two rows) and (2) a post-onset-weighting mechanism followed by the symmetric double-exponential temporal window used by Bernstein *et al.* (2001) (lower two rows).

Task	a	b	T_a	T_b
ITD	3.1	−2.1	2.8 ms	5.9 ms
IID	1.8	−0.8	1.8 ms	13.0 ms
ITD	1.1	−0.9	2.0 ms	20.3 ms
IID	1.8	−1.9	4.6 ms	11.6 ms

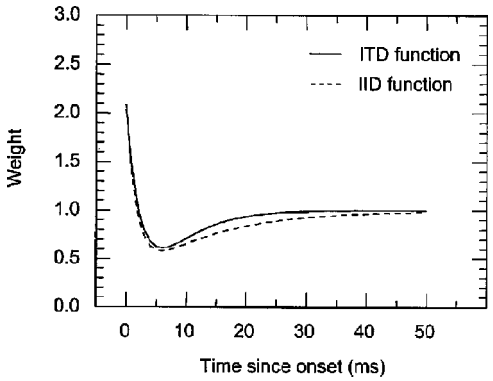


FIG. 6. Illustration of the post-onset-weighting functions derived in Sec. III B. The solid line shows the function derived for the threshold ITDs and the dashed line shows the function derived from the threshold IIDs.

In an effort to evaluate the generality of our approach, we sought to determine how well the combination of a post-onset-weighting mechanism and a temporal window could account for the measurements obtained by Zurek (1980), which served as the motivation for the present study. Figure 7 displays the threshold IIDs obtained by Zurek averaged across his two listeners (symbols).[2] Two sets of predicted values are shown. The dashed line represents predictions derived when the values of the parameters of the post-onset-weighting function and the parameters of the temporal window were fixed and equal to those derived for our listeners (see Tables II and III). For this analysis, only W_0 (the effective IID) was free to vary, and its value was adjusted via an iterative procedure in order to maximize the amount of variance accounted for. The solid line represents predictions derived when all parameters were free to vary and their optimal values found using the method described in Sec. III B.

The predictions derived by using the values of the window parameters derived for our listeners (dashed-line) accounted for 69% of the variance in the threshold IIDs obtained by Zurek (1980). The fit underestimates the observed thresholds for forward fringes of 5 and 10 ms. Allowing the values of the parameters of the post-onset and temporal window functions to vary (solid line) results in a much improved and excellent fit to the data that accounts for 98% of the variance in the observed thresholds. The equations used to compute these predictions are shown to the right of the plot. The values of the parameters are such that, in order to account for Zurek's thresholds, the post-onset weighting mechanism must incorporate a larger relative weighting of the binaural information between 0 and 5 ms than was the case for our listeners. The reasons for this discrepancy are unclear; they may come about as a result of inter-individual differences. Nevertheless, our analyses demonstrate that the combination of a post-onset-weighting mechanism with a temporal window can provide highly successful accounts of both our data and those obtained by Zurek (1980).

C. An alternative form of the temporal window

In this third analysis, we report the degree to which the use of the complex temporal window derived by Bernstein *et al.* (2001) can, in conjunction with a post-onset-weighting function, account for the data in the current study. Bernstein *et al.* (2001) employed an experimental paradigm similar to the one used in the current study but in which the probe segment was always temporally centered within a longer burst of noise that was, in separate conditions, either diotic or interaurally uncorrelated. The durations of the probe segment and of the surrounding noise were varied independently. Bernstein *et al.* reported that a *single* temporal window could account for both threshold ITDs and IIDs across all of their conditions. Their ''double exponential'' window was symmetric, being formed from the sum of two exponential functions:

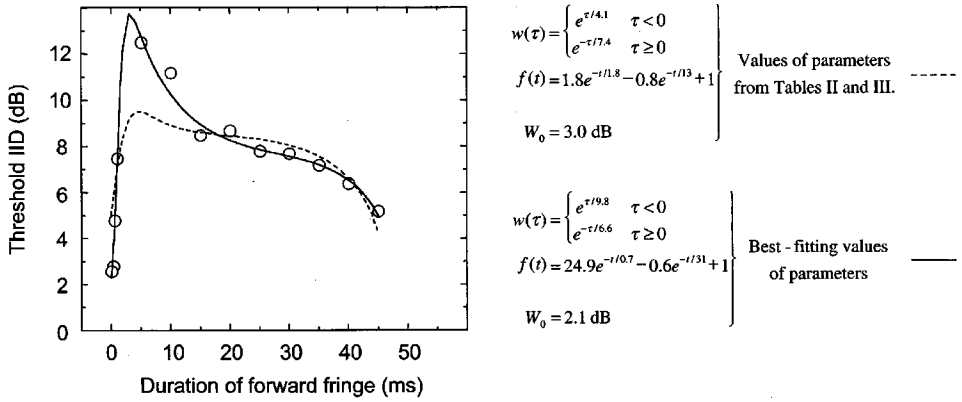


FIG. 7. IIDs (in dB) as a function of the duration of the forward fringe. The symbols represent the thresholds obtained by Zurek (1980) averaged across his two listeners. The dashed line represents predictions derived when the values of the parameters of the post-onset-weighting function and the parameters of the temporal window were fixed and equal to those derived for our listeners (see Tables II and III). The solid line represents predictions derived when the values of all parameters were free to vary as described in Sec. III B.

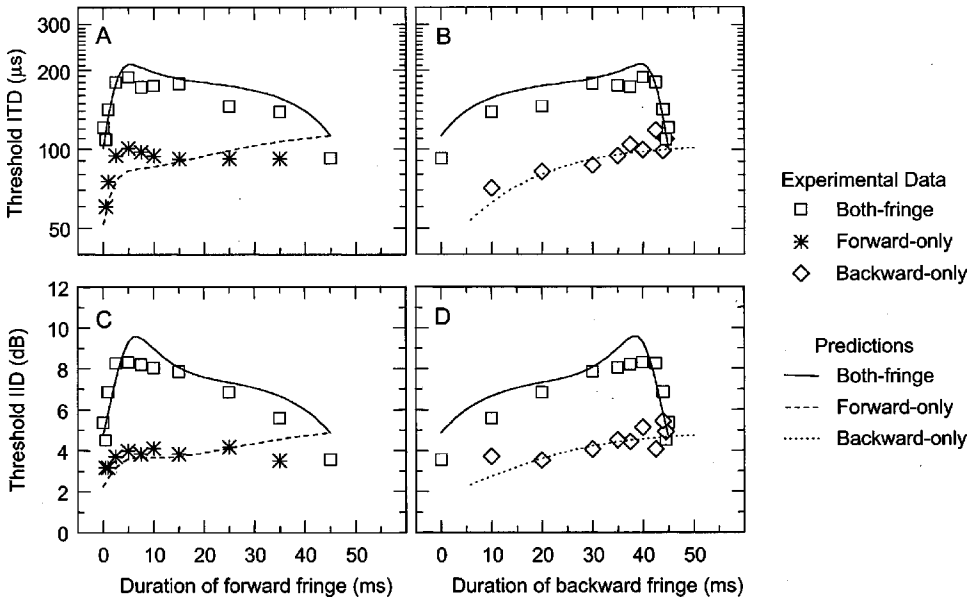


FIG. 8. Same as Fig. 5 with the exception that the predictions were derived using a post-onset-weighting function in combination with the symmetric temporal window from Bernstein *et al.* (2001).

$$w(\tau) = 0.94e^{(-|\tau|/0.084)} + (1-0.94)e^{(-|\tau|/13.8)}, \quad (5)$$

where τ (in ms) represents time relative to the peak response of the window. As was discussed by Bernstein *et al.*, this window places great emphasis on information within its temporal center, and far less, albeit important, emphasis on the temporally surrounding binaural information.

In order to test the generality of this window function, we considered how well it, in conjunction with an appropriate post-onset-weighting function, could account for the data in the present study. Specifically, we repeated the fitting procedure described in Sec. III B except that the parameters of the temporal window were *fixed* and equal to those in Eq. (5). That is, only W_0 and the values of the parameters of the post-onset-weighting function were free to vary.

The results of this analysis are depicted in Fig. 8 in the same format as that used in Fig. 5. Overall, the fits to the data are remarkable, considering that the temporal window function employed was taken from a different, albeit related, study. The post-onset-weighting functions derived in conjunction with the fixed temporal window were quite similar to those derived earlier in Sec. III B (see rows 3 and 4 of Table III). The values differed in that, in conjunction with the temporal window function of Bernstein *et al.* (2001), the weight at onset was somewhat less for ITDs, and the duration of the region of ''suppressed weight'' was slightly longer. We believe that the similarity of the derived post-onset-weighting functions across the two different analyses attests to the validity of utilizing such a function.

The predictions derived from this analysis account for 83% of the variance in both the threshold ITDs and threshold IIDs. The fits are somewhat poorer than those reported in Sec. III B (Fig. 5) in that they overestimate the thresholds in the both-fringe conditions for durations of the forward fringe in the region of 5 to 10 ms. We interpret the overall success of these predictions to indicate that the symmetric, double-exponential window reported by Bernstein *et al.* (2001) is also consistent with the data in this study.

IV. DISCUSSION

We measured thresholds for the detection of interaural disparities conveyed by a 5-ms burst of broadband noise (termed the probe segment) that was embedded within a longer, diotic noise. The paradigm was similar to one employed by Zurek (1980). Two major outcomes were observed. First, we found that thresholds were influenced by the duration of the diotic noise that occurred *both* before and after the probe segment. Second, like Zurek, we observed a rapid elevation in thresholds when the onset of the probe segment occurred between 1 and 10 ms after the onset of the diotic noise.

The first outcome is consistent with the notion of a binaural temporal window that acts to average the interaural disparities within the probe segment with those within the temporally surrounding noise. The concept of a temporal window is supported by a variety of previous investigations (e.g., Grantham and Wightman, 1978; Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998; Holube *et al.*, 1998; Akeroyd and Summerfield, 1998, 1999). The general form of the data in the present study were accounted for, quantitatively, by such a mechanism. The time constants of the window functions derived in this study were on the order of 5 ms, yielding equivalent rectangular durations of about 10 ms. These values, like those obtained by Bernstein *et al.* (2001), are substantially shorter than those often derived from studies characterizing the binaural system as ''sluggish'' (e.g., Grantham and Wightman, 1978, 1979; Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998; Holube *et al.*, 1998; Akeroyd and Summerfield, 1998, 1999). The time constants reported in those studies typically range from about 50 to 250 ms, depending on the experimental paradigm and on the

individual listener. We concur (as did Bernstein *et al.*) with Kollmeier and Gilkey (1990) who suggested that different tasks may tap different aspects of binaural temporal processing, and thus yield different time constants.

Despite accounting for the overall trends in the data, a temporal window, by itself, could not capture the second of the two major outcomes mentioned above. That is, it could not account for the rapid rise in thresholds as the onset of the probe segment was delayed between 1 and 10 ms from the onset of the diotic noise. We were able to account for this aspect of the data by invoking the operation of a post-onset-weighting function *in combination with* a temporal window.

The form of the post-onset-weighting function, as derived from our data, was characterized by a region of ''suppressed weight'' shortly after its onset (Houtgast and Aoki, 1994). That is, as suggested by Zurek (1980), it appears that the leading portion of the noise acts, in some manner, to briefly reduce sensitivity to subsequent binaural information. In our view, it is not necessarily the case that this reduced sensitivity is the result of some type of physiologically based inhibition. Rather, the logical possibility exists that the relative insensitivity to binaural cues that occurs shortly after the onset of stimulation is a manifestation of peripheral interactions and/or more central mechanisms that effectively ''weigh'' differentially binaural disparities across time.

It should be noted that we offer the present model as an account of how sensitivity to binaural information is affected when the interaural cues within a brief portion of a longer sound are altered. As such, we wish to emphasize that we do not consider the present model to be a general model of the precedence effect and related phenomena. The quantitative accounts of the precedence effect that have been offered (see Litovsky *et al.*, 1999) are either descriptive of behavioral measures (Shinn-Cunningham *et al.*, 1993, 1995),[3] or include mathematical specifications of binaural mechanisms or processes that are used to derive predictions of thresholds (e.g., Lindemann, 1986; Tollin, 1998). These models have been used to account for the results of experiments in which listeners were presented with a pair of clicks or brief bursts of noise. To our knowledge, none have been extended to account for sensitivity to binaural information within a brief portion of a longer sound. The model presented here which is, in effect, a quantitative implementation of a conceptual model of the precedence effect described by Zurek (1987) provides a successful account of such measures.

ACKNOWLEDGMENTS

We thank Drs. Tino Trahiotis and Klaus Hartung for their comments on earlier versions of this report. We also thank the two anonymous reviewers and the Associate Editor (Dr. Sid Bacon) for their helpful suggestions and criticisms. The work was supported by research Grants nos. NIH DO-04073 and NIH DC-00234 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health. The work was also supported by a Medical Research Council Career Development Fellowship awarded to the first author. A portion of this work was reported at the 139th meeting of the Acoustical Society of America, Atlanta, GA [M. A. Akeroyd, L. R. Bernstein, and C. Trahiotis, J. Acoust. Soc. Am. **107**, 2822 (2000)].

[1]The formula used to compute the percentage of the variance for which our predicted values of threshold accounted was

$$100\times\left(1-\frac{\Sigma(O_i-P_i)^2}{\Sigma(O_i-\overline{O})^2}\right),$$

where O_i and P_i represent the individual observed and predicted values of threshold, respectively, and $\overline{O}$ represents the mean of the observed values of threshold.

[2]A similar analysis could not be performed for the threshold ITDs measured by Zurek because, for both of his listeners, several thresholds exceeded the maximum ITD available.

[3]In the study by Shinn-Cunningham *et al.* (1993), behavioral measures are assumed to be based upon an ''effective'' internal ITD, α. The value of α is derived via a weighted sum of the external ITDs, ITD_1 and ITD_2, imposed on the first click and second click of a pair of brief stimuli, respectively. This is expressed as $\alpha=c\times\mathrm{ITD}_1+(1-c)\times\mathrm{ITD}_2$, where the value of c ranges between 0 and 1. The predicted thresholds derived with the model described in the present report are also based on a weighted sum of ITDs, although it is the weighted sum of the ITDs present *throughout* a burst of noise rather than of the ITDs of two successive stimuli. That both models use a weighted sum suggests that it would be interesting to determine whether the values of c computed by Shinn-Cunningham *et al.* can be derived quantitatively from the present model, which incorporates a post-onset-weighting mechanism in combination with a temporal window.

Akeroyd M. A., and Summerfield, Q. (**1998**). ''Predictions of signal thresholds in frozen-noise masker using monaural and binaural windows,'' in *Psychological and Physiological Advances in Hearing*, edited by A. Palmer, A. Rees, Q. Summerfield, and R. Meddis (Whurr, London, 1998).

Akeroyd, M. A., and Summerfield, A. Q. (**1999**). ''A binaural analog of gap detection,'' J. Acoust. Soc. Am. **105**, 2807–2820.

ANSI (**1969**). ANSI S4.6-1969, ''Specifications for audiometers'' (American National Standards Institute, New York).

Aoki, A., and Houtgast, T. (**1992**). ''A precedence effect in the perception of inter-aural cross-correlation,'' Hear. Res. **59**, 25–30.

Bernstein, L. R., Trahiotis, C., Akeroyd, M. A., and Hartung, K. (**2001**). ''Sensitivity to brief changes of interaural time and interaural intensity,'' J. Acoust. Soc. Am. **109**, 1604–1615.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge).

Culling, J. F., and Summerfield, Q. (**1998**). ''Measurements of the binaural temporal window using a detection task,'' J. Acoust. Soc. Am. **103**, 3540–3553.

Freyman, R. L., Clifton, R. K., and Litovsky, R. Y. (**1991**). ''Dynamic processes in the precedence effect,'' J. Acoust. Soc. Am. **90**, 874–884.

Grantham, D. W., and Wightman, F. L. (**1978**). ''Detectability of varying interaural temporal differences,'' J. Acoust. Soc. Am. **63**, 511–523.

Grantham, D. W., and Wightman, F. L. (**1979**). ''Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation,'' J. Acoust. Soc. Am. **65**, 1509–1517.

Holube, I., Kinkel, M., and Kollmeier, B. (**1998**). ''Binaural and monaural auditory filter bandwidths and time constants on probe-tone detection experiments,'' J. Acoust. Soc. Am. **104**, 2412–2425.

Houtgast, T., and Aoki, S. (**1994**). ''Stimulus-onset dominance in the perception of binaural information,'' Hear. Res. **72**, 29–36.

Kollmeier, B., and Gilkey, R. H. (**1990**). ''Binaural forward and backward masking: Evidence for sluggishness in binaural detection,'' J. Acoust. Soc. Am. **87**, 1709–1719.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Lindemann, W. (**1986**). ''Extension of a binaural cross-correlation model by contralateral inhibition,'' J. Acoust. Soc. Am. **80**, 1623–1630.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). ''The precedence effect,'' J. Acoust. Soc. Am. **106**, 1633–1654.

Saberi, K. (**1995**). ''Some considerations on the use of adaptive methods for estimating interaural-delay thresholds,'' J. Acoust. Soc. Am. **98**, 1803–1806.

Shinn Cunningham, B. G., Zurek, P. M., and Durlach, N. I. (**1993**). ''Adjustment and discrimination measurements of the precedence effect,'' J. Acoust. Soc. Am. **93**, 2923–2932.
Shinn-Cunningham, B. G., Zurek, P. M., Durlach, N. I., and Clifton, R. K. (**1995**). ''Cross-frequency interactions in the precedence effect,'' J. Acoust. Soc. Am. **98**, 164–171.
Tollin, D. J. (**1998**). ''Computational model of the lateralization of clicks and their echoes,'' in *Proceedings of the NATO Advanced Study Institute on Computational Hearing*, Lucca, Italy, 1–12 July 1998, edited by S. Greenberg and M. Slaney, pp. 77–82.
Tollin, D. J., and Henning, G. B. (**1998**). ''Some aspects of the lateralization of echoed sound in man. I. The classical interaural-delay based precedence effect,'' J. Acoust. Soc. Am. **104**, 3030–3038.
Zurek, P. M. (**1980**). ''The precedence effect and its possible role in the avoidance of interaural ambiguities,'' J. Acoust. Soc. Am. **67**, 952–964.
Zurek, P. M. (**1987**). ''The precedence effect,'' in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York).

Informational and energetic masking effects in the perception of multiple simultaneous talkers

Douglas S. Brungart[a)]
Air Force Research Laboratory, Human Effectiveness Directorate, 2610 Seventh Street, Wright-Patterson AFB, Ohio 45433

Brian D. Simpson
Veridian, 5200 Springfield Pike, Suite 200, Dayton, Ohio 45431

Mark A. Ericson and Kimberly R. Scott
Air Force Research Laboratory, Human Effectiveness Directorate, 2610 Seventh Street, Wright-Patterson AFB, Ohio 45433

(Received 13 February 2001; accepted for publication 13 August 2001)

Although many researchers have examined the role that binaural cues play in the perception of spatially separated speech signals, relatively little is known about the cues that listeners use to segregate competing speech messages in a monaural or diotic stimulus. This series of experiments examined how variations in the relative levels and voice characteristics of the target and masking talkers influence a listener's ability to extract information from a target phrase in a 3-talker or 4-talker diotic stimulus. Performance in this speech perception task decreased systematically when the level of the target talker was reduced relative to the masking talkers. Performance also generally decreased when the target and masking talkers had similar voice characteristics: the target phrase was most intelligible when the target and masking phrases were spoken by different-sex talkers, and least intelligible when the target and masking phrases were spoken by the same talker. However, when the target-to-masker ratio was less than 3 dB, overall performance was usually lower with one different-sex masker than with all same-sex maskers. In most of the conditions tested, the listeners performed better when they were exposed to the characteristics of the target voice prior to the presentation of the stimulus. The results of these experiments demonstrate how monaural factors may play an important role in the segregation of speech signals in multitalker environments.
 [DOI: 10.1121/1.1408946]

PACS numbers: 43.66.Pn, 43.66.Rq, 43.71.Gv [LRB]

I. INTRODUCTION

Many everyday listening situations require the extraction of information from speech signals that are masked by one or more simultaneous competing talkers. In most of these situations, the competing speech signals originate from different locations in the room and the listeners can take advantage of differing inputs to the two ears to spatially segregate the competing messages. This is the classic ''cocktail party'' problem that was first described by Cherry (1953) and has been extensively studied over the past 50 years [see Bronkhorst (2000) and Ericson and McKinley (1997) for recent reviews of the literature in this area]. However, when the target and masking speech originate from the same direction relative to the listener, or when the competing speech signals are presented monaurally or diotically, no binaural segregation cues are available and the listeners must rely on monaural cues to segregate the competing messages. Examples of monaural speech segregation cues include differences in the individual vocal characteristics of the target and masking talkers [vocal tract size, fundamental frequency ($F0$), accent, speaking style, etc.], differences in the prosodic features of the target and masking speech, and differences in the overall levels of the target and masking signals (Darwin and Hukin, 2000; Bregman, 1994).

Previous studies that have examined the diotic or monaural perception of two competing speech signals have shown that differences in the vocal characteristics of the competing talkers, such as differences in target and masker sex, can dramatically improve the intelligibility of the target speech (Brungart, 2001b; Festen and Plomp, 1990). These studies have also shown that listeners can use differences in the levels of the two talkers to selectively attend to the quieter talker in the stimulus (Brungart, 2001b), and that signal-to-noise ratio (SNR) consequently has relatively little influence on the intelligibility of the target talker at SNRs from 0 dB to −10 dB (Egan, Carterette, and Thwing, 1954; Dirks and Bower, 1969).

The results of these 2-talker experiments suggest that differences in the vocal characteristics and the overall levels of the competing talkers would be important in the perception of 3-talker or 4-talker stimuli, but they do not provide any direct quantitative evidence to support this hypothesis. Previous studies that have directly examined the perception of three or more simultaneous talkers have focused primarily on binaural segregation cues (the cocktail-party phenomenon) and not on monaural segregation of speech

a)Author to whom correspondence should be addressed; electronic mail: douglas.brungart@wpafb.af.mil

(Abouchacra *et al.*, 1997; Ericson and McKinley, 1997; Crispien and Ehrenberg, 1995; Drullman and Bronkhorst, 2000; Hawley *et al.*, 1999; Nelson *et al.*, 1999; Peissig and Kollmeier, 1997; Yost *et al.*, 1996). Of the handful of experiments that have examined the perception of three or more monaurally or diotically presented speech signals, the majority have done so only indirectly, either as a control condition in a cocktail-party experiment (Hawley *et al.*, 2000; Drullman and Bronkhorst, 2000; Yost *et al.*, 1996; Ericson and McKinley, 1997) or as a control condition in a dichotic listening experiment (Carhart *et al.*, 1969). Only one early experiment (Miller, 1947) systematically examined the effects of varying both the SNR and the number of competing talkers on the perception of a diotic multitalker stimulus. These studies have shown that the intelligibility of the target talker decreases when additional competing talkers are added to the stimulus and when the level of the target speech is reduced relative to the levels of the competing talkers in the stimulus.

One important factor that has not yet been fully explored is the effect that different configurations of target and masker sex have on the perception of multitalker speech stimuli. Although previous experiments with two talkers have shown that voice characteristics play an important role in speech segregation (Brungart, 2001b), very little effort has been made to systematically examine the effects of target and masker sex in listening environments with more than two competing talkers.

This paper describes a series of experiments examining the effects that differences in the vocal characteristics and overall levels of the competing talkers have on the perception of a target phrase in a multitalker speech signal. These experiments were based on a previous experiment that used the coordinate response measure (CRM) to examine the effects of target and masker sex and target-to-masker ratio (TMR) on the perception of two simultaneous talkers (Brungart, 2001b). The CRM method, which requires listeners to identify one of eight numbers and one of four colors in each target phrase, was selected both to allow direct comparison with the earlier 2-talker results and to emphasize the effects of informational masking in the multitalker results. Note that the term ''informational masking'' refers to listening situations where the target and masker signals are clearly audible but the listener is unable to segregate the elements of the target signal from the elements of the similar-sounding distracters (Freyman *et al.*, 1999; Doll and Hanna, 1997; Kidd *et al.*, 1995; Kidd *et al.*, 1994; Watson *et al.*, 1976). This differs from traditional ''energetic masking,'' where competing signals overlap in time and frequency in such a way that portions of one or more of the signals are rendered inaudible. The results of an earlier 2-talker experiment (Brungart, 2001b) showed that speech perception with the CRM was dominated by informational masking: the listeners were generally able to hear both competing speech messages, but they had difficulty segregating the content of the target phrase from the content of the masking phrase. Although it is reasonable to expect the effects of energetic masking to increase when more talkers are added to the stimulus, the strong informational masking effects found in that 2-talker experiment suggest that informational masking may also influence performance in the CRM task when the target phrase is masked by more than one competing talker.

Three different experiments were conducted. The first experiment, which was essentially a 3-talker and 4-talker extension of an earlier 2-talker experiment (Brungart, 2001b), examined the effects of TMR and target and masker voice characteristics when all of the masking voices were presented at the same level relative to the target phrase. The second experiment examined 3-talker listening situations where all three talkers were presented at different levels. The third experiment examined how providing *a priori* information about the vocal characteristics of the target talker affected performance in the CRM task.

II. EXPERIMENT 1: EFFECTS OF TARGET AND MASKER SEX AND TARGET-TO-MASKER RATIO

A. Methods

1. Stimuli

The target and masking phrases used in this experiment were taken directly from the publicly available CRM speech corpus for multitalker communications research (Bolia *et al.*, 2000). This corpus, which is based on a speech intelligibility test first developed by Moore (1981), consists of phrases of the form ''Ready (call sign) go to (color) (number) now'' spoken with all possible combinations of eight call signs (''Arrow,'' ''Baron,'' ''Charlie,'' ''Eagle,'' ''Hopper,'' ''Laker,'' ''Ringo,'' ''Tiger''); four colors (''blue,'' ''green,'' ''red,'' ''white''); and eight numbers (1–8). Thus, a typical utterance in the corpus would be ''Ready Baron go to blue five now.'' Eight talkers (four male, four female) were used to record each of the 256 possible phrases, so a total of 2048 phrases are available in the corpus.

In the speech-on-speech masking conditions of this experiment, each stimulus presentation consisted of three or four simultaneous phrases from the CRM corpus: a target phrase with the call sign ''Baron'' and two or three masker phrases with different randomly selected call signs other than ''Baron.'' The CRM phrases in each trial were randomly selected with the restriction that all of the target and masker phrases contained different color coordinates and different number coordinates.

The overall level (RMS power) of each masker phrase was set to the same level (approximately 60–70 dB SPL). The overall level (RMS power) of the target phrase was adjusted relative to the levels of the masker phrases to produce 1 of 10 randomly chosen TMRs ranging from −12 dB to +15 dB in 3 dB steps. The target and masker phrases were added together, and the combined signal was randomly roved over a 6 dB range (in 1 dB steps) before being presented to the listener diotically over headphones. This roving prevented the listeners from using level to differentiate between the target and masking talkers, and ensured that the results were not dependent on any one particular overall stimulus level. Note that throughout this paper target-to-masker ratio (TMR) refers to the ratio of the target talker to one individual masking talker, while signal-to-noise ratio (SNR) refers to the ratio of the target talker to the total combined masking signal. Thus, when the levels of all three talkers are the same

TABLE I. Number of trials collected in each experimental condition.

Condition	Total trials	Trials per TMR
TD[a]	8828	883
TS[a]	6884	688
TT[a]	2288	289
TM[a]	2400	480
TN[a]	9020	902
TDD	3542	354
TSD	4792	479
TSS	2825	283
TTT	2161	216
TMM	2500	357
TDDD	2053	205
TSDD	2208	221
TSSD	2053	205
TSSS	2196	220
TTTT	2277	228
TMMM	2520	360

[a]Data collected in previous 2-talker experiment (Brungart, 2001b).

in the 3-talker condition of this experiment, the TMR of each masking talker would be 0 dB and the overall SNR would be approximately −3 dB.

The different voice configurations used in the experiment are referred to by alphabetic codes describing the relative similarities of the target and masking voices. The first letter in the code is always a T, representing the target phrase. Additional letters are added to represent each masking phrase in the stimulus: a T represents a phrase spoken by the same talker used in the target phrase; an S represents a different talker of the same sex as the target talker; and a D represents a talker who was different in sex than the target talker. For example, in the 3-talker conditions, four different target and masker voice configurations were tested (Table I). In the TDD configuration, the two masking talkers were both different in sex than the target talker. In the TSD configuration, one of the masking talkers was the same sex as the target talker, and the other masking talker was a different sex than the target talker. In the TSS configuration, the masking phrases were both spoken by talkers who were the same sex as the target talker. And in the TTT configuration, the same talker was used in the target phrase and in both masking phrases.

In addition to these speech-on-speech masking conditions, two additional conditions were tested in which the masking talkers were replaced by two (TMM) or three (TMMM) envelope-modulated speech-shaped noise maskers. The speech-shaped noise used for these masking signals was produced by spectrally shaping Gaussian noise with a FIR filter matching the average long-term RMS spectrum of the 2048 sentences in the CRM corpus (Brungart, 2001b). This spectrally shaped noise was then modulated with the envelopes of two or three randomly selected competing speech phrases from the CRM corpus (calculated by convolving the absolute value of the speech waveform with a 7.2 ms rectangular window) and added together in order to produce a noise signal with roughly the same temporal distribution of energy as the 2- or 3-talker competing speech signals. The level of each modulated noise masker was determined from the overall RMS power of the noise waveform, and set to the same level used for each competing talker in the speech-on-speech masking conditions. The modulated noise conditions were tested at 7 different target-to-masker ratios ranging from −12 dB to 6 dB in 3 dB steps.

2. Listeners

Nine paid listeners, five male and four female, participated in the experiment. All had normal hearing (15 dB HL from 500 Hz to 6 kHz) and their ages ranged from 21 to 55. Each had participated in previous auditory experiments, and all had previous experience in an earlier experiment using the same speech materials with two competing talkers (Brungart, 2001b).

3. Procedure

The listeners participated in the experiment while seated at a control computer in a quiet listening room. On each trial, the speech stimulus was generated by a sound card in the control computer (Soundblaster AWE-64) and presented to the listener diotically over headphones (Sennheiser HD-520). Then an eight-column, four-row array of colored digits corresponding to the response set of the CRM was displayed on the CRT, and the listener used the mouse to select the colored digit corresponding to the color and number used in the target phrase containing the call sign ''Baron.''

The trials in the speech-on-speech masking conditions were divided into blocks of 120, each taking approximately 12 minutes to complete. Within each block of trials, each talker in the corpus was used as the speaker of the target phrase in exactly 15 trials. The masking talkers were selected randomly in order to produce a roughly even balance across the different possible configurations of target sex and masker sex shown in Table I. The masking talker selection in the 3-talker condition was initially done randomly, which produced a large number of instances of the TSD configuration (roughly 48%) and only a small number of instances of the TTT configuration (roughly 5%). In order to produce a more even distribution across the talker configurations in the overall results, the selection process was changed approximately 2/3 of the way into the data collection process to select the TTT configuration in 40% of the trials and the TSD configuration in 10% of the trials. This manipulation resulted in the overall distribution of trials shown in Table I. On each trial of the 4-talker configuration, the target and maskers were equally likely to occur in any of the five talker-masker configurations shown in Table I. All other variables in the 3- and 4-talker presentations, including the masking call signs, the numbers and colors of the target and masker phrases, and the target-to-masker ratio, were chosen randomly with replacement on each trial. Each of the nine listeners first participated in approximately 2000 trials in the 3-talker condition, resulting in 1480 valid trials per subject which were used in the subsequent analysis.[1] Each listener then participated in 1200 trials in the 4-talker condition.

The modulated-noise trials were collected after the speech-on-speech masking trials were completed. The trials in each of the modulated noise conditions were divided into

two blocks of 140 trials collected on different days. The TMM configuration was collected first, followed by the TMMM configuration. The total number of trials collected in each condition, as well as the number of trials for each target-to-masker ratio tested in each condition, is shown in Table I.

4. Comparison of 3-talker and 4-talker results to 2-talker results

In order to provide additional insight into the 3-talker and 4-talker conditions examined in this experiment, the results of these conditions are compared directly to the results of an earlier experiment that examined 2-talker speech-on-speech masking with the same nine listeners and the same CRM speech materials used in this experiment (Brungart, 2001b). A total of three 2-talker configurations were tested in that experiment (denoted by the superscript ''a'' in Table I): a different-sex masking condition (TD), a same-sex masking condition (TS), and a same talker masking condition (TT). In addition, a TM configuration was tested with a single modulated speech-spectrum-shaped noise masker of the same type used in the TMM and TMMM configurations of this experiment, and a TN configuration was tested in which a continuous speech-spectrum-shaped noise masker was rectangularly gated to the same length as the target phrase. Because all of the procedures used in these 2-talker conditions were essentially the same as those used in the 3- and 4-talker conditions of this experiment, the results of the two experiments are directly comparable.

B. Results and discussion

Figure 1 shows the percentage of trials where the listener correctly identified both the color and the number in the target phrase as a function of target-to-masker ratio for each of the different target-masker configurations in the experiment. The 2-talker configurations are shown in the top panel of the figure, the 3-talker configurations are shown in the middle panel of the figure, and the 4-talker configurations are shown in the bottom panel of the figure. The data have been averaged across the nine listeners used in the experiment, and the error bars in the figure represent the 95% confidence interval of each data point. The results in the figure indicate that overall performance depended on both the relative similarities between the target and masking voices and the TMR. This was verified by conducting three separate within-subject ANOVAs on the factors of TMR and target-masker configuration for the 2-talker, 3-talker, and 4-talker data shown in each panel of Fig. 1. In these analyses, the percentages of correct responses were calculated separately for each listener for each combination of target-masker configuration and TMR. These percentages were then transformed with the arcsine transform and used as the dependent variables in the within-subject ANOVAs. The results of these analyses showed that the main effect of TMR and target-masker configuration and the interactions between TMR and target-masker configuration were significant at the $p < 0.001$ level for each of the three panels of Fig. 1.

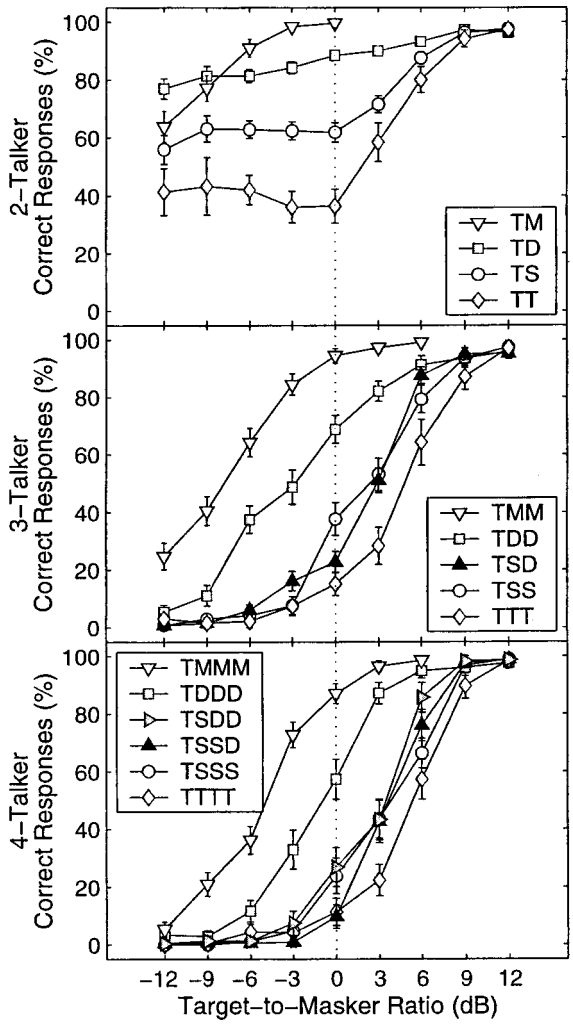


FIG. 1. Percentage of trials in Experiment 1 in which the listeners correctly identified both the color and number coordinates in the target phrase as a function of the target-to-masker ratio. The top panel shows results from an earlier 2-talker experiment that used the same speech materials and listeners as the 3-talker and 4-talker conditions tested in this experiment (Brungart, 2001a, 2001b). The middle panel shows results for the 3-talker conditions tested in this experiment, and the bottom panel shows results for the 4-talker conditions tested in this experiment. The legends indicate the different target-masker voice configurations used: T=target voice; S=voice of different talker of the same sex as the target voice; D=voice of a talker of the opposite sex of the target voice; M=envelope-modulated noise masker. The error bars represent 95% confidence intervals (±1.96 standard errors) in each condition.

The effects of the different voice configurations were relatively consistent across the 2-, 3-, and 4-talker conditions. In general, performance was best when the target voice was qualitatively different than the masking voices, and performance was worse when the target voice was qualitatively similar to the masking voices. The percentage of correct responses was consistently highest when the target talker was a different sex than any of the masking talkers (TD, TDD, and TDDD, shown by the squares in the figure) and the percentage of correct responses was generally lowest when the same voice was used in the target and masking phrases (TT, TTT, and TTTT, shown by the diamonds in the figure). Performance in the other voice configurations varied systematically with the qualitative similarity of the target and masking voices only when the target phrase was presented at a sub-

stantially higher level than the masking phrases. At a TMR of +6 dB, for example, performance decreased systematically with the number of masking talkers who were the same sex as the target talker in both the 3-talker configurations (TDD>TSD>TSS) and the 4-talker configurations (TDDD>TSDD>TSSD>TSSS). When the TMR was reduced to +3 dB, however, the performance differences between the same-sex and mixed-sex masking configurations were drastically reduced (TSD≈TSS; TSSS ≈TSDD≈TSSD) and, when the TMR was reduced to 0 dB (indicated by the vertical dotted line in the figure), performance was actually substantially lower in some of the mixed-sex masking configurations than in the same-sex masking configurations. Specifically, in those 3- and 4-talker configurations where one masking talker was a different sex than the other talkers in the stimulus (TSD, TSSD), performance was lower than in the corresponding same-sex masking configurations (TSS, TSSS) when the TMR was near 0 dB. This effect, which we refer to as ''odd-sex distraction,'' cannot be explained by traditional theories of energetic or informational masking. Energetic masking would be expected to cause more interference with a same-sex masker than with a different-sex masker because of the greater spectral overlap of two same-sex speech signals. One might also expect more informational masking to occur when the masker is qualitatively similar to the target than when the masker is qualitatively different from the target. Odd-sex distraction appears to be a special form of informational masking in which a particularly salient masker causes the listener's attention to be drawn away from the target phrase. This occurs only when the overall levels of the talkers in the stimulus are similar enough (TMRs near 0 dB) that the listener must rely entirely on vocal characteristics to segregate the competing talkers. The strength of this distracting effect is illustrated by the number of trials where the listeners' responses matched both the color and number spoken by the odd-sex masking talker. In the TSD configuration, the listeners were nearly as likely to respond with both the color and number spoken by the odd-sex masking talker as they were to respond with both the color and number spoken by the target talker (21% vs 24% at a TMR of 0 dB). In the TSSD configuration, they were actually more likely to respond with the coordinates spoken by the odd-sex talker than with the coordinates spoken by the target talker (17% versus 10%). Odd-sex distraction is examined in more detail in Experiments 2 and 3.

Further insights into the differences between the 2-talker, 3-talker, and 4-talker conditions can be obtained by examining the results as a function of the overall decreases in SNR that occur when additional masking talkers are added to the stimulus. At a TMR of 0 dB, where all the talkers are speaking at the same level, the total power present in the masking speech (relative to the target speech) is 0 dB in the 2-talker condition, approximately 3 dB in the 3-talker condition, and approximately 4.8 dB in the 4-talker condition. Figure 2 plots performance in the different-sex (top row), same-sex (second row), same-talker (third row), and noise-masker (bottom row) conditions of the experiment as a function of TMR (left column) and as a function of SNR (right column). As in Fig. 1, the data have been averaged across the 9 listeners in the experiment and the error bars represent 95% confidence intervals for each data point.

The TMR data plotted in the left column of the figure show that, within a given voice configuration, there was a substantially larger degradation in performance when the number of talkers was increased from two to three than when the number of talkers was increased from three to four (Fig. 2, left column). This occurred because overall performance degraded substantially more with decreasing TMR in the 3-talker and 4-talker conditions than in the 2-talker condition. In the 2-talker condition, performance was essentially independent of TMR at TMR values lower than 0 dB. In the 3-talker and 4-talker conditions, however, performance decreased monotonically with decreasing TMR in every target-masker configuration tested. Other multitalker experiments have also shown that the addition of a second interfering talker produces a much larger decrease in performance than the addition of the initial interfering talker (Hawley *et al.*, 2000; Drullman and Bronkhorst, 2000; Miller, 1947).

The SNR data plotted in the right column of the figure show that the number of talkers present in the stimulus had different effects on performance at positive and negative SNRs. At positive SNRs, performance generally increased with the number of competing talkers. This may have occurred because the difference in level between the target talker and any one of the masking talkers (TMR) increased as the number of competing talkers increased at a fixed SNR. If the listeners were using differences in the overall levels of the talkers as a means to segregate the target speech from the masking speech, this could explain why they performed better as the number of competing talkers increased at a fixed SNR.

At negative SNRs, the number of talkers had a somewhat different effect on the intelligibility of the target phrase. The listeners were apparently able to focus their attention on the less intense voice in the 2-talker stimulus even at SNRs as low as −12 dB. However, in the 3- and 4-talker stimuli, there is no indication that they were able to use differences in level to segregate the less intense target speech from the more intense masking speech. Performance in the 3–4 talker conditions dropped off rapidly with decreasing SNR, and was in all cases near chance when the SNR was less than −9 dB (Fig. 2, right panels). It is interesting to note that performance at negative SNRs was almost identical in the 3-talker and 4-talker conditions in each of the three speech-on-speech voice configurations shown in Fig. 2. This suggests that performance at negative SNRs with three or more competing talkers depends primarily on the voice characteristics and total power of the masking speech, and not on the number of masking talkers in the stimulus.

The results from the modulated noise masking conditions indicate that energetic masking was probably not a controlling factor in any of the listening conditions tested. Performance in the conditions with two or three modulated noise maskers (triangles in Fig. 1) was substantially better than in any of the conditions with a corresponding number of speech maskers. The modulated noise maskers contained roughly the same temporal distribution of energy as the

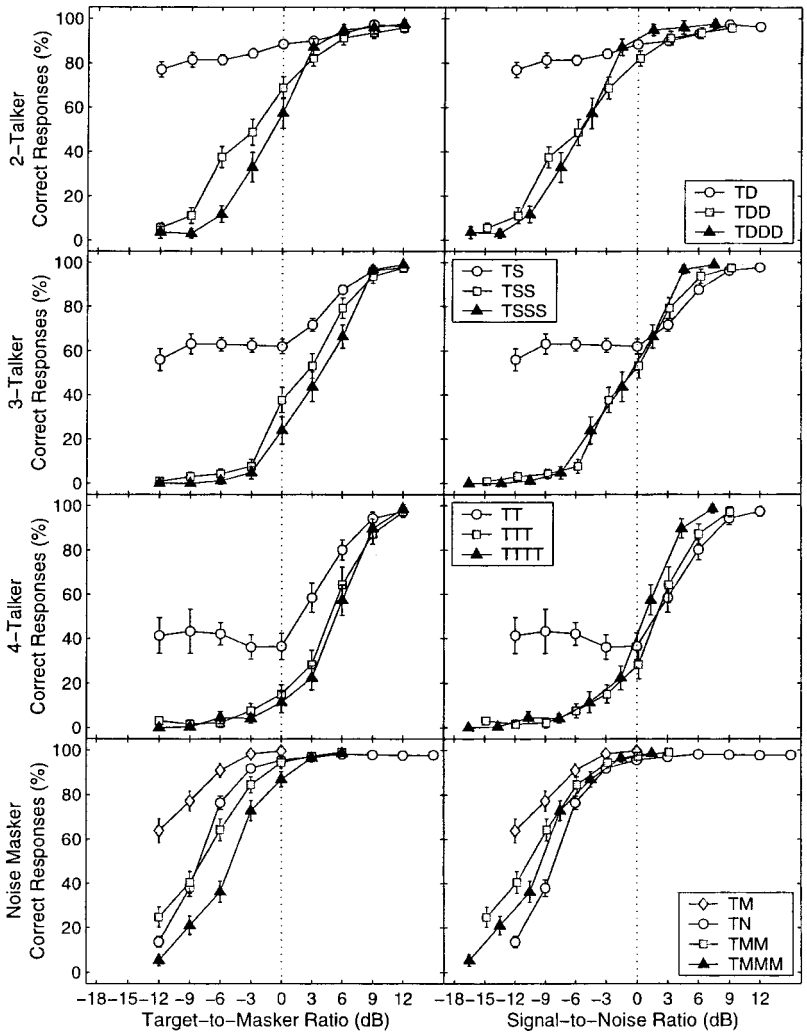


FIG. 2. Correct color and number identifications in the 2-talker, 3-talker, and 4-talker conditions of Experiment 1 for different-sex maskers (top row), same-sex maskers (second row), same-talker maskers (third row), and modulated noise maskers (bottom row). The left column shows the results as a function of target-to-masker ratio, which is defined as the level of the target voice relative to the levels of the individual masker voices. The right column shows the results as a function of SNR, which is the level of the target voice (RMS power) relative to the summed level of all of the masker voices (RMS power). The error bars represent 95% confidence intervals (± 1.96 standard errors) in each condition.

speech signals, but their performance curves were consistently shifted 3 dB to the left of the easiest speech masker configurations tested (TDD, TDDD), and as much as 12–15 dB to the left of the more difficult speech masker configurations (TTT, TTTT). In other words, the modulated noise signal had to be 3–15 dB more intense (in terms of overall RMS power) than a multitalker speech signal to produce the same amount of masking. Although the time-frequency energy distributions in the modulated noise maskers were not identical to those in the speech maskers, it does not seem likely that they were different enough to account for these differences in masking effectiveness. Thus, there is reason to believe that some form of nonenergetic masking was occurring even in the performance curves at negative SNRs in the 3-talker and 4-talker conditions. The roles of informational and energetic masking in multitalker speech perception are further explored in Experiments 2 and 3.

When the number of modulated noise talkers was increased at a fixed SNR, overall performance systematically decreased (Fig. 2, bottom right panel). It is likely that this occurred because the temporal distribution of energy was more uniform when the number of modulated noise maskers was increased and the listeners were less able to listen to the target phrase between ''the gaps'' of the masking signal (Bronkhorst and Plomp, 1992). The performance levels in the TMM and TMMM configurations fell between those for the single modulated noise masking condition (TM) and the continuous (unmodulated) noise masking conditions (TN). This is essentially the same pattern of performance reported by Bronkhorst and Plomp (1992) for speech intelligibility as the number of speech-modulated noise maskers increased.

III. EXPERIMENT 2: MASKING EFFECTS OF TWO COMPETING TALKERS WITH DIFFERENT TMRs

In the first experiment, all of the masking voices in the stimulus were presented at the same level relative to the target speech. In many real-world listening situations, however, the target and masking voices are all at different levels. In order to determine how intelligibility is affected by differences in the TMRs of the different masking talkers, a second experiment was conducted in which the levels of two masking talkers were varied independently. In half of the trials, both masking talkers were the same sex as the target; in the remaining trials, one masking talker was the same sex as the target and the other was a different sex.

A. Methods

The procedures used in the second experiment were nearly identical to those used in the first experiment. The

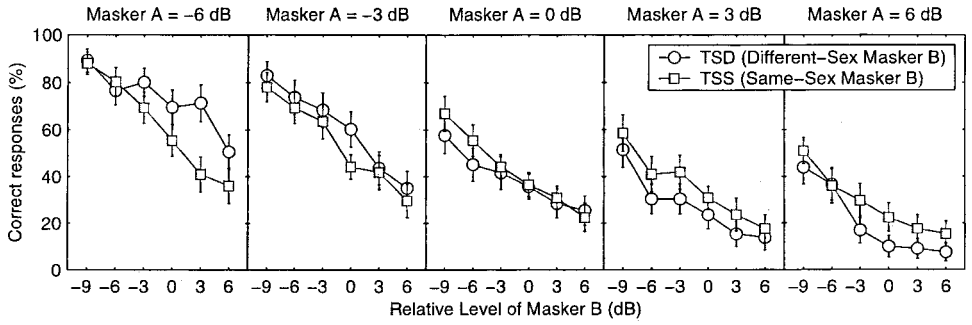


FIG. 3. Percentage of correct color and number identifications in Experiment 2 as a function of the relative levels of a same-sex masker phrase (Masker A) and a same- or different-sex masking phrase (Masker B). The five panels represent different relative levels of Masker A. In the leftmost panel, Masker A was presented at a level 6 dB lower than the target talker, and in the rightmost panel Masker A was presented at a level 6 dB higher than the target phrase. Within each panel, each symbol represents a different relative level of Masker B. The two curves in each panel represent same-sex and different-sex voices for Masker B, as indicated in the legend. The error bars represent 95% confidence intervals (±1.96 standard errors) in each condition.

speech signals were selected randomly from the CRM speech corpus, adjusted to the correct TMR according to the RMS powers of the target and masking speech signals, electronically combined, output through a DAC (Tucker–Davis DD1), and presented to the listeners diotically over headphones. The listeners then responded with the color and number coordinates contained in the target phrase addressed to the call sign ''Baron.''

The target-masker configurations used in the second experiment were somewhat different than those used in the first experiment. Every stimulus presentation contained three different talkers: the target phrase; a same-sex masking phrase (Masker A) presented at a level (relative to the target) of −6 dB, −3 dB, 0 dB, 3 dB, or 6 dB; and a same-sex or different-sex masking phrase (Masker B) presented at a level (relative to the target) of −9 dB, −6 dB, −3 dB, 0 dB, 3 dB, or 6 dB. In all, a total of 45 unique target-masker configurations were tested in the second experiment.

Eight paid volunteer listeners (3 males, 5 females) participated in the experiment. All eight were also participants in the first experiment. Each listener participated in a total of 1200 trials, divided into six blocks of 180 trials each plus one additional block of 120 trials. Within each block, the trials were randomly distributed across the 45 different possible target-masker configurations. The total experiment was completed over a two-week period, with each listener participating in 1-2 blocks of trials each day.

B. Results and discussion

Figure 3 shows the percentage of correct color and number identifications as function of the relative level of Masker B for each of the five different relative levels of Masker A tested in Experiment 2. The results have been averaged across the eight listeners used in the experiment, and they are plotted separately for the conditions with a same-sex Masker B (circles) and those with a different-sex Masker B (squares). These results show that the percentage of correct responses systematically decreased when the relative level of either of the two masking voices was increased. This can be seen from the decrease in the percentage of correct responses that occurred when the relative level of Masker A was increased at a fixed level of Masker B (moving from the left panel to the right panel of Fig. 3) and when the relative level of Masker B was increased at a fixed level of Masker A (moving from left to right within each panel of Fig. 3). A three-factor within-subject ANOVA conducted on the arcsine-transformed individual listener results for the factors of Masker A TMR, Masker B TMR, and Masker B Sex verified that the main effects of Masker A TMR and Masker B TMR were both significant at the $p<0.001$ level.

In general, performance was degraded more by the addition of a same-sex Masker B when the level of the same-sex Masker A was lower than the level of the target voice (left panel of Fig. 3), and was degraded more by the addition of a different-sex Masker B when the level of the same-sex Masker A was higher than the level of the target voice (right panel of Fig. 3). The three-factor ANOVA conducted on the results showed that this interaction between Masker A TMR and Masker B sex was significant at the $p<0.001$ level ($F_{(4,420)}=3.454$). Traditional theories of energetic masking, based on the spectral overlap of the target and masking signals, and informational masking, based on the qualitative similarity of the target and masking signals, would predict better performance in the TSD condition at all levels of Masker A. The superior performance that occurred in the TSS condition when Masker A was 6 dB more intense than the target phrase seems to be directly related to the odd-sex distraction that occurred at TMRs near 0 dB in Experiment 1. Apparently listeners are more susceptible to distraction from an odd-sex talker than to interference from a same-sex talker when they are listening to the quieter of two same-sex voices. This increased vulnerability to distraction may be related to the intense concentration required to selectively focus attention on the quieter of two same-sex talkers, which is inherently more difficult than listening to the more intense talker in the stimulus. Apparently the presence of a more salient different-sex voice makes it difficult to remain focused on the quieter of two same-sex talkers.

It is not clear why the performance advantage of the TSS configuration over the TSD configuration did not extend to the case where all three talkers were presented at the same level (center panel of Fig. 3), as it did in Experiment 1. One possible explanation is that the listeners were able to learn that the target phrase was never spoken by the odd-sex talker in the TSD configuration of the second experiment (as it was in the TDD configuration of the first experiment) and that

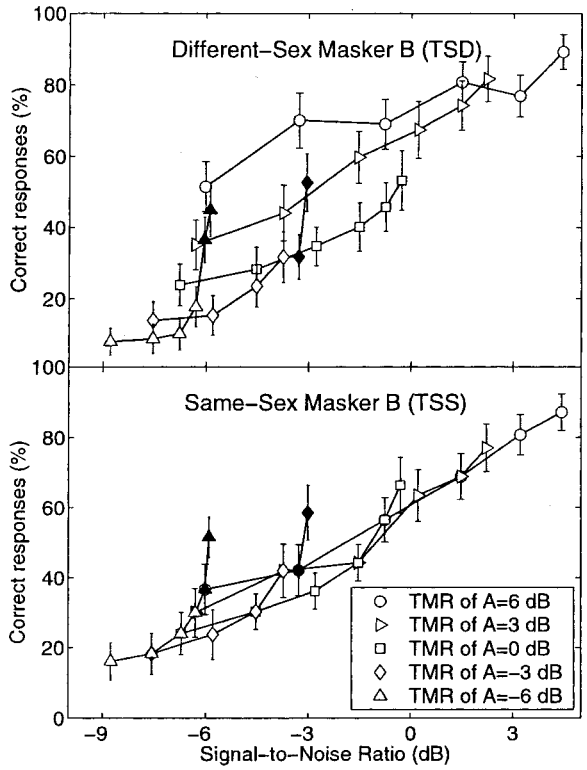


FIG. 4. Performance in Experiment 2 plotted as a function of SNR. The top panel shows the data from the TSD configuration, and the bottom panel shows the data from the TSS configuration. Within each panel, the different curves represent nine different levels of Masker B (relative to the target) at the fixed level of Masker A (relative to the target) shown in the legend. The filled symbols represent points where Masker A was presented at a level that was 12 dB or more higher than the level of Masker B (see text for details). The error bars represent 95% confidence intervals (± 1.96 standard errors) in each condition.

they adopted a strategy of selectively ignoring the odd-sex talkers in the TSD configurations of Experiment 2. The results of a third experiment, which is discussed later, will show that listeners are able to selectively ignore the odd-sex talker when they have *a priori* knowledge about the sex of the target talker.

The results shown in Fig. 3 are replotted in Fig. 4 as a function of overall SNR. In other words, each symbol in Fig. 4 represents one of the symbols shown in Fig. 3 plotted as a function of the ratio of the RMS energy in the target speech to the RMS energy in the combined two-talker masking signal. Different symbols have been used for each level of Masker A, and the data for the TSS and TSD conditions have been plotted in separate panels. In both the TSS and TSD configurations, the resulting performance curves increased abruptly with respect to the SNR whenever Masker A was presented at a level that was 12 dB or more higher than the level of Masker B (filled symbols in Fig. 4). It is likely that these abrupt increases occurred because the combined signal from the target talker and Masker A was intense enough to render Masker B inaudible, effectively reducing the TSS and TSD listening tasks to the much easier 2-talker TS listening task. Thus, the curves in these configurations may reflect a transition from the relatively low level of performance that occurs in 3-talker listening at negative SNRs to the relatively high level of performance that occurs in 2-talker listening at negative SNRs (see Fig. 2). The large improvements in performance that occurred with almost negligible increases in SNR in these curves provide strong evidence that informational masking plays an important role in 3-talker speech perception—energetic masking alone would not be expected to generate such large changes in performance with such small changes in SNR. These curves also suggest that our ability to attend to the quieter talker in a 2-talker stimulus is ''fragile'': performance in this task appears to drop off precipitously as soon as a third talker becomes audible, even when that talker has a negligible effect on overall SNR of the stimulus.

In conditions where there was no abrupt increase in performance with respect to SNR (open symbols in Fig. 4), there was a clear distinction between the performance curves for the TSD configuration shown in the top panel of Fig. 4 and the performance curves for the TSS configuration shown in the bottom panel of Fig. 4. In the TSD configuration, the vertical positions decrease systematically as the relative level of Masker A was increased. These curves show that performance in the TSD configuration depended on both the individual levels of the masking talkers and the overall SNR of the stimulus. However, the curves in the lower panel of Fig. 4 suggest that performance in the TSS configuration depended almost exclusively on the overall SNR of the stimulus. The individual levels of the masking talkers had little impact on the results, and the data from the TSS configuration were clustered into a tight distribution, suggesting an almost linear relation between overall SNR and the percentage of correct identifications (linear correlation coefficient $r = 0.98$). A seemingly related result was found in the 3- and 4-talker same-sex configurations of Experiment 1. Specifically, a given SNR led to similar performance for both the 3-talker and 4-talker same-sex configurations (see Fig. 2, second panel, right column). These results suggest that masking signals consisting of two or more same-sex talkers are effectively grouped together into a single same-sex, multiple-talker masking signal. When the overall SNR of the stimulus is less than 0 dB, the impact of this multitalker masking noise on performance is primarily determined by its total power and not by the number or levels of its component talkers.

IV. EXPERIMENT 3: SELECTIVE AND DIVIDED ATTENTION WITH 2, 3, OR 4 COMPETING TALKERS

When listening to multiple competing speech messages, a distinction must be made between divided-attention tasks, where the listeners must simultaneously monitor all the speech channels for pertinent information that might come from any of the competing talkers, and selective-attention tasks, where the listeners know *a priori* which talker they should listen for and they are attempting to focus their attention on the target talker while ignoring the masking talkers (Yost, 1997; Yost *et al.*, 1996; Abouchacra *et al.*, 1997). The CRM-based speech perception test that was used in Experi-

ments 1 and 2 is, in effect, a combination of these two types of tasks: initially, the listeners must divide their attention across all of the competing talkers to determine which one is directly addressing the call sign ''Baron;'' then, they must selectively attend to this target talker in order to extract the color and number coordinates from the target phrase. It is, however, possible to change this CRM listening task into a selective-attention task by providing the listeners with *a priori* information about the vocal characteristics of the target talker. In order to examine the effects of selective and divided attention on multitalker speech perception, a third experiment was conducted that compared performance in a task where the listeners knew the vocal characteristics of the target talker *a priori* (the ''selective-attention'' condition) and a task where the listeners had no knowledge about the identity of the target talker prior to hearing the stimulus (divided attention plus selective attention; the ''divided-attention'' condition).

A. Methods

The procedures used in the third experiment were very similar to the procedures used in the first and second experiments. The target and masker phrases in each trial were randomly selected from the CRM corpus, equalized, electronically summed, and presented to the listeners diotically over headphones. The listeners then identified the color and number coordinates used in the target phrase containing the call sign ''Baron'' by moving the mouse to the appropriate colored number on the CRT of the control computer.

Six paid volunteer listeners participated in the third experiment. All six had previously participated in Experiment 1, and five of the six were also participants in Experiment 2.

A total of nine different target-masker configurations were used in the third experiment: two 2-talker configurations (TS and TD); three 3-talker configurations (TSS, TSD, and TDD); and four 4-talker configurations (TSSS, TSSD, TSDD, and TDDD). In all of these configurations, the overall RMS power of each competing talker was normalized to the same level as the RMS power of the target talker (TMR=0 dB). The target-masker configurations were randomly selected (with replacement) prior to each trial of the experiment. Note that the divided attention conditions replicated a subset of the conditions collected in Experiment 1 when the target-to-masker ratio was 0 dB.

The experiment was divided into two different conditions. In the divided-attention condition, the target talkers in each block of 180 trials were selected randomly from the eight talkers in the corpus. In the selective-attention condition, the same target talker was used in all of the 180 trials in each block, and, prior to beginning data collection in each block, 10 training trials were provided in which only the target talker was presented. This enabled the listeners to become familiar with the characteristics of the target voice. Each listener participated in eight blocks of 180 trials in the selective-attention condition (one for each of the eight talkers in the corpus), and four blocks of 180 trials in the divided-attention condition. The ordering of the two conditions (and of the different talkers in the selective-attention conditions) was randomized across the six different listeners used in the experiment.

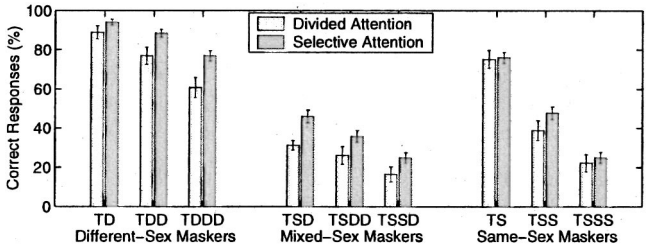


FIG. 5. Percentage of correct color and number identifications in the selective and divided attention conditions of Experiment 3. All data were collected with a TMR of 0 dB. The error bars represent 95% confidence intervals (±1.96 standard errors) in each condition.

B. Results

The results of the experiment are shown in Fig. 5. Each pair of bars compares performance in the divided attention task to performance in the selective attention task for a different target-masker configuration. The data have been grouped separately for configurations with different-sex maskers, configurations with mixed-sex maskers, and the configurations with same-sex maskers.

In general, performance improved when the listeners were provided with *a priori* information about the characteristics of the target voice. A two-factor within-subject ANOVA on the arcsine-transformed individual data for the factors of *a priori* information and target-masker configuration showed that the main effect of *a priori* exposure to the target voice was significant at the $p<0.001$ level ($F_{(1,90)}=32.278$). This improvement was, on average, substantially larger in the different-sex and mixed-sex masking configurations (+12% and +10%, respectively, averaged across all the different-sex and mixed-sex configurations) than in the same-sex masking configurations (+4%). The improvement was also substantially larger in the 3-talker and 4-talker configurations (+12% and +9%, respectively) than in the 2-talker configurations (+4%). The only two voice configurations in which *a priori* information did not produce a substantial improvement in performance were the same-sex 2-talker and 4-talker conditions (TS and TSSS). The relative ineffectiveness of the *a priori* voice information in these same-sex conditions can be explained by a closer examination of the distribution of incorrect responses in the experiment. The data plotted in Fig. 6 show the proportion of trials in which the listeners incorrectly responded with both the color and number coordinates spoken by one of the masking talkers in the experiment. The top panel shows the proportion of responses that matched one of the masking talkers of a different sex than the target talker, and the bottom panel shows the proportion of responses that matched one of the masking talkers of the same sex as the target talker. These data show that the *a priori* voice information provided in the selective attention conditions led to a substantial reduction in the number of different-sex confusions, but had no meaningful effect on the number of same-sex confusions. This result indicates that most of the useful information the listeners were able to obtain from *a priori* exposure to the target talker

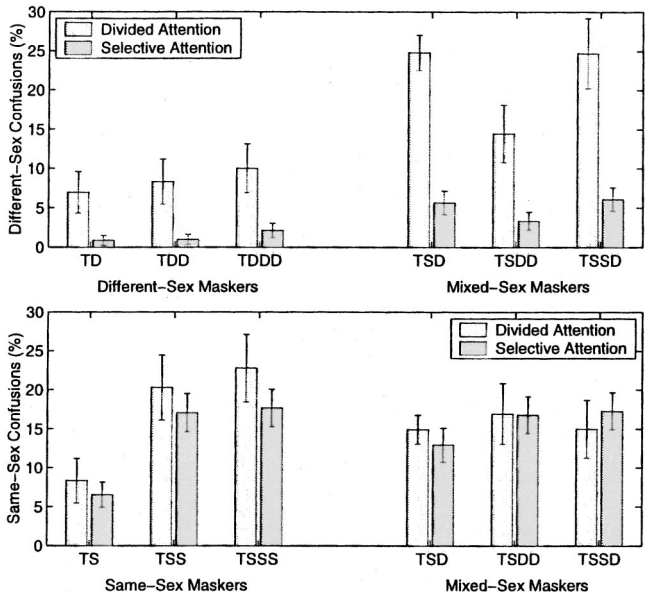


FIG. 6. Same-sex confusions and different-sex confusions in Experiment 3. Same-sex confusions occurred when the listener's responses included both the color and number spoken by one of the same-sex masking talkers in the stimulus. Different-sex confusions occurred when the listeners' responses included both the color and number spoken by one of the different-sex masking talkers in the stimulus. The error bars represent 95% confidence intervals (± 1.96 standard errors) in each condition.

was the sex of the target talker. They were not very good at exploiting the subtle variations in the voices of the same-sex talkers used in the experiment. This would explain why the performance improvements in the selective attention condition were much smaller in the same-sex configurations than in different-sex and mixed-sex configurations.

There is also a relatively simple explanation for why the *a priori* voice information produced less improvement in the 2-talker conditions than in the 3- and 4-talker conditions. When a multitalker speech stimulus contains only two talkers, listeners really only need to attend to one of the two call signs in order to successfully identify the color and number in the target phrase. If they happen to initially focus their attention on the correct talker, they can simply maintain their focus on that talker for the remainder of the carrier phrase. If they happen to initially focus their attention on the wrong talker, then by elimination they know that the target call sign was spoken by the unattended talker and they simply need to switch their attention to the other voice for the remainder of the carrier phrase. Alternatively, listeners may be able to attend to the incorrect voice for the entire stimulus and then, realizing it was the wrong voice, retrieve the coordinates of the correct voice from their short-term auditory memories. In either case, the use of the process of elimination to reduce the 2-talker divided-attention task into the much easier selective-attention task could explain why performance with two talkers was not much different in the selective and divided-attention conditions.

One important difference between the selective- and divided-attention conditions is that the *a priori* voice information provided in the selective-attention condition essentially eliminated the odd-sex distraction that occurred in Experiment 1 and in the divided-attention conditions of Experiment 3. In the divided-attention condition, performance was significantly worse in the TSD configuration than in the TSS configuration, and significantly worse in the TSSD configuration than in the TSSS configuration ($p < 0.05$, two-tailed *t*-tests). In the TSSD configuration, odd-sex distraction was so strong that the listeners were significantly more likely to respond with the color and number spoken by the odd-sex masking talker (25%) than with the color and number spoken in the target phrase (16%) ($p < 0.05$, two-tailed *t*-test). In the selective-attention condition, however, there was no meaningful difference between the TSSS and TSSD configurations, or between the TSS and TSD configurations. Odd-sex distraction is apparently only a factor when the listener is forced to simultaneously monitor all of the competing call signs to find the target call sign, and it does not occur when the listeners know they should ignore the odd-sex talker prior to being exposed to the stimulus. This may indicate that odd-sex distraction occurs because the listeners tend to initially focus their attention on the more salient odd-sex talker in the stimulus, which frequently causes them to miss the information presented in the target phrase. When they know *a priori* that they should not listen to the odd-sex talker, they perform no worse in the odd-sex masking configurations than in the same-sex masking configurations.

The distribution of incorrect responses in Experiment 3 can also be used to evaluate the influence of energetic masking in multitalker listening. Mutual energetic masking between the competing talkers in the stimulus should be greatest when all of the simultaneous talkers are speaking at the same level, as they were in Experiment 3. If this mutual energetic masking were powerful enough to make the competing speech messages inaudible, one would expect the distribution of incorrect responses to be random across all of the different incorrect colors and numbers in the response set. In this experiment, however, the incorrect responses were not randomly distributed: virtually all of the listeners' color and number responses were present in one of the masking phrases in the stimulus. The percentage of responses containing either a color or a number not contained in the stimulus never exceeded 4% in any of the configurations tested in Experiment 3. These percentages are almost negligible in comparison to the percentages of randomly selected responses one would expect to contain a color or number not in the stimulus, which would range from 50% in the 4-talker configurations to more than 85% in the 2-talker configurations. This suggests that at least some, and perhaps all, of the competing voices were audible in the multitalker stimuli of this experiment when the TMR was 0 dB, and that most of the incorrect responses resulted from the listeners' inability to segregate the target phrase from the masking phrases (informational masking) and not from the listeners' inability to hear the target phrase (energetic masking).

V. SUMMARY AND CONCLUSIONS

The results of this study may help to reveal some of the strategies listeners use to process monaural speech signals

with two or more competing talkers. One of the most important findings is that there are fundamental differences in the multitalker speech perception task when the target talker is more intense than the masking talkers and when one or more of the masking talkers is more intense than the target talker.

Target level higher than masker level: When the target-to-masker ratio is positive, performance is generally best when the target voice is qualitatively different than the masking voices. Different-sex maskers degrade performance less than same-sex maskers, and same-sex maskers degrade performance less than same-talker maskers. Performance also generally increases with the number of competing talkers when the overall SNR of the stimulus is fixed.

Target level at or below masker level: When the target-to-masker ratio is negative, performance is much worse when there are three or four competing talkers than when there are only two competing talkers. Overall performance is generally less dependent on target-masker similarity than it is at positive TMRs: performance is still best in the different-sex conditions, but there is little difference between the mixed-sex and same-sex masking conditions, and, in some cases, listeners can be severely distracted by a single masking talker that is different in sex than the other talkers in the stimulus. This odd-sex distraction effect is strongest when all of the competing talkers are presented at the same level, and probably occurs because listeners have a natural tendency to initially focus their attention on the most salient talker in a stimulus. When the maskers are all the same sex as the talker, performance depends almost exclusively on the SNR, and not on the number or individual levels of the masking talkers.

When the SNR of the stimulus is fixed and the listeners are given *a priori* information about the voice characteristics of the target talker, performance generally improves only in mixed-sex and different-sex listening conditions. Most of this improvement results from a reduction in the number of responses matching the different-sex masking voices in the stimulus. The number of same-sex confusions is unaffected by the *a priori* exposure to the target voice. Prior information about the voice characteristics of the target does, however, effectively eliminate odd-sex distraction.

There is also evidence that most of these effects result from informational masking rather than energetic masking. Performance in these experiments was substantially worse in the speech-masking conditions than in the conditions with a corresponding number of modulated noise maskers. In addition, only a small number of the incorrect responses at a TMR of 0 dB contained words which were not present in the stimulus. Although energetic masking certainly played some role at the lowest TMRs used in this experiment, in most cases it appears that performance was most influenced by informational masking. In light of these results, it is important to note that the particular speech test used in these experiments (the CRM) is likely to be substantially more sensitive to informational masking than other types of speech tests that have been used in previous multitalker experiments. At least four factors contribute to this sensitivity: (1) In contrast to most other multitalker speech tasks, which use a target talker that is easily discriminated from the masking talkers on the basis of vocal characteristics (Festen and Plomp, 1990; Drullman and Bronkhorst, 2000), location (Drullman and Bronkhorst, 2000; Hawley *et al.*, 1999; Crispien and Ehrenberg, 1995), or onset time (Freyman *et al.*, 1999), the CRM speech task used in Experiments 1 and 2 of this study uses a randomly selected target talker that is identifiable only by the use of the call sign ''Baron'' in the target phrase. This introduces a substantial amount of uncertainty about the identity of the target talker both in the call sign portion of the carrier phrase and in the color-number portion of the carrier phrase. (2) Although the listeners know the positions of the call sign and coordinate words within the CRM phrases, the phrases themselves contain no contextual information. Thus, all three critical words (call sign, color, and number) must be correctly identified to correctly respond in the CRM talk. This is in direct contrast to natural speech, which contains contextual clues that can be used to reconstruct sentences even when some of the words in the sentences are unintelligible. (3) All of the critical words in the CRM are aligned to occur almost simultaneously, which would almost never happen in natural speech. (4) The key color and number words in the CRM are drawn from small vocabularies of relatively dissimilar words. This allows listeners to correctly guess the right color or number even when they are only able to hear a small portion of the phonetic information in the word. For example, the color ''green'' could be distinguished from the other colors in the CRM corpus either by the initial consonant sound ⟨gr⟩, the vowel ⟨ee⟩, or the final consonant sound ⟨n⟩. Thus, listeners are able to correctly identify the color green from any one of these three phonetic components even if the other two are obscured by energetic masking. This phonetic redundancy makes it possible to measure informational masking effects with the CRM in listening configurations where the target speech would be rendered completely unintelligible by energetic masking in many other speech perception tests. Because the CRM is so highly tuned to measuring informational masking, the results of this experiment should be viewed not as a general indicator of the role informational masking plays in all multitalker speech tasks, but rather as an isolated study of the effects of informational masking in a context where energetic masking is relatively unimportant.

Perhaps the most intriguing aspect of these findings is the fragility of our ability to attend to the quieter of two simultaneous talkers. It is apparent from these results that the mechanisms humans use to focus their attention on the quieter of two talkers are severely disrupted as soon as a third talker becomes audible in the stimulus. Further research is needed to determine how listeners focus their attention on the quieter talker in a 2-talker signal and why they are no longer able to do so when a third talker is added to the stimulus. A thorough understanding of this process might facilitate the development of algorithms for unmixing speech signals at negative SNRs that would be tremendously beneficial in the development of advanced hearing aids and speech recognition systems.

ACKNOWLEDGMENTS

The authors would like to thank Chris Darwin, Adelbert Bronkhorst, and Kim Abouchacra for their helpful comments in the paper. This work was supported by AFOSR Grant No. 01-HE-01-COR.

[1]The other trials were collected early in the experiment and contained invalid configurations including trials where the target talker was used for one of the two masking talkers and trials where more than one talker used the same color or number coordinates. These trials were eliminated from the data analysis.

Abouchacra, K., Tran, T., Besing, J., and Koehnke, J. (**1997**). ''Performance on a selective attention task as a function of stimulus presentation mode,'' Proceedings of the Midwinter Meeting of the Association for Research in Otolaryngology.

Bolia, R., Nelson, W., Ericson, M., and Simpson, B. (**2000**). ''A speech corpus for multitalker communications research,'' J. Acoust. Soc. Am. **107**, 1065–1066.

Bregman, A. S. (**1994**). *Auditory Scene Analysis* (MIT, Cambridge).

Bronkhorst, A. (**2000**). ''The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,'' Acustica **86**, 117–128.

Bronkhorst, A., and Plomp, R. (**1992**). ''Effects of multiple speechlike maskers on binaural speech recognition in normal and impaired listening,'' J. Acoust. Soc. Am. **92**, 3132–3139.

Brungart, D. (**2001a**). ''Evaluation of speech intelligibility with the coordinate response measure,'' J. Acoust. Soc. Am. **109**, 2276–2279.

Brungart, D. (**2001b**). ''Informational and energetic masking effects in the perception of two simultaneous talkers,'' J. Acoust. Soc. Am. **109**, 1101–1109.

Carhart, R., Tillman, T., and Greetis, E. (**1969**). ''Perceptual masking in multiple sound backgrounds,'' J. Acoust. Soc. Am. **45**, 694–703.

Cherry, E. (**1953**). ''Some experiments on the recognition of speech, with one and two ears,'' J. Acoust. Soc. Am. **25**, 975–979.

Crispien, K., and Ehrenberg, T. (**1995**). ''Evaluation of the 'Cocktail Party Effect' for multiple speech stimuli within a spatial audio display,'' J. Audio Eng. Soc. **43**, 932–940.

Darwin, C., and Hukin, R. (**2000**). ''Effectiveness of spatial cues, prosody, and talker characteristics in selective attention,'' J. Acoust. Soc. Am. **107**, 970–977.

Dirks, D., and Bower, D. (**1969**). ''Masking effects of speech competing messages,'' J. Speech Hear. Res. **12**, 229–245.

Doll, T., and Hanna, T. (**1997**). ''Directional cueing effects in auditory recognition,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, NJ).

Drullman, R., and Bronkhorst, A. (**2000**). ''Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation,'' J. Acoust. Soc. Am. **107**, 2224–2235.

Egan, J., Carterette, E., and Thwing, E. (**1954**). ''Factors affecting multichannel listening,'' J. Acoust. Soc. Am. **26**, 774–782.

Ericson, M., and McKinley, R. (**1997**). ''The intelligibility of multiple talkers spatially separated in noise,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, NJ), pp. 701–724.

Festen, J., and Plomp, R. (**1990**). ''Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing,'' J. Acoust. Soc. Am. **88**, 1725–1736.

Freyman, R., Helfer, K., McCall, D., and Clifton, R. (**1999**). ''The role of perceived spatial separation in the unmasking of speech,'' J. Acoust. Soc. Am. **106**, 3578–3587.

Hawley, M., Litovsky, R., and Culling, J. (**2000**). ''The 'cocktail party' effect with four kinds of maskers: Speech, time-reversed speech, speech-shaped noise, or modulated speech-shaped noise,'' Proceedings of the Midwinter Meeting of the Association for Research in Otolaryngology, p. 31.

Hawley, M., Litovsky, R., and Colburn, H. (**1999**). ''Speech intelligibility and localization in a multi-source environment,'' J. Acoust. Soc. Am. **105**, 3436–3448.

Kidd, G. J., Mason, C., Deliwala, P., Woods, W., and Colburn, H. (**1994**). ''Reducing informational masking by sound segregation,'' J. Acoust. Soc. Am. **95**, 3475–3480.

Kidd, G. J., Mason, C., and Rohtla, T. (**1995**). ''Binaural advantage for sound pattern identification,'' J. Acoust. Soc. Am. **98**, 1977–1986.

Miller, G. (**1947**). ''Sensitivity to changes in the intensity of white Gaussian noise and its relation to masking and loudness,'' J. Acoust. Soc. Am. **191**, 609–619.

Moore, T. (**1981**). ''Voice communication jamming research,'' *AGARD Conference Proceedings 331: Aural Communication in Aviation* (Neuilly-Sur-Seine, France), pp. 2:1–2:6.

Nelson, W. T., Bolia, R. S., Ericson, M. A., and McKinley, R. L. (**1999**). ''Spatial audio displays for speech communication. A comparison of free-field and virtual sources,'' Proceedings of the 43rd Meeting of the Human Factors and Ergonomics Society, pp. 1202–1205.

Peissig, J., and Kollmeier, B. (**1997**). ''Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners,'' J. Acoust. Soc. Am. **35**, 1660–1670.

Watson, C., Kelly, W., and Wroton, H. (**1976**). ''Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty,'' J. Acoust. Soc. Am. **60**, 1176–1185.

Yost, W. (**1997**). ''The cocktail party problem: Forty years later,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, NJ), pp. 329–348.

Yost, W., Dye, R., and Sheft, S. (**1996**). ''A simulated 'cocktail party' with up to three sources,'' Percept. Psychophys. **58**, 1026–1036.

High-speed digital imaging of the medial surface of the vocal folds

David A. Berry[a)]
Department of Biomedical Engineering and Department of Speech Pathology & Audiology, The University of Iowa, Iowa City, Iowa 52242

Douglas W. Montequin
Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242

Niro Tayama
Department of Otolaryngology, University of Tokyo, Tokyo, Japan

(Received 2 April 2001; accepted for publication 15 August 2001)

High-speed digital imaging of the medial surface of the vocal folds was performed in excised canine larynx experiments. Building on the excised larynx investigations of Baer [Ph.D. dissertation, MIT, Boston, MA (1975)] and hemilarynx investigations of Jiang and Titze [Laryngoscope **103**, 872–882 (1993)], nine vocal fold fleshpoints were tracked simultaneously along the medial surface of one coronal plane of the left vocal fold using a Kodak EktaPro 4540 high-speed digital imaging system. By imaging from two distinct views, 3D reconstructions of fleshpoint trajectories were performed with a sampling frequency of 4.5 kHz and a spatial resolution of approximately 0.08 mm. Quantitative results were derived from a typical example of periodic chestlike vibrations. Furthermore, these data were decomposed into empirical eigenfunctions, the building blocks of vocal fold vibration, illuminating basic mechanisms of self-sustained oscillation. Previously, such mechanisms have only been explored theoretically using computer models of vocal fold vibration [Berry *et al.*, J. Acoust. Soc. Am. **95**, 3595–3604 (1994)]. Similar to the theoretical studies, two eigenfunctions captured 98% of the variance of the data. Because this investigation utilized high-speed technology, the methodology may also be used to examine complex, aperiodic vibrations. Thus, this technique allows mechanisms of regular and irregular vocal fold vibration to be explored using direct observations of vibrating tissues in the laboratory. [DOI: 10.1121/1.1408947]

PACS numbers: 43.70.Aj [AL]

I. INTRODUCTION

For years, it has been noted that octave jumps, low-frequency modulations, and irregular vibrations are characteristic of signals from disordered voices (Koike, 1969; Sirviö and Michelsson, 1976; Wieser, 1981; Ludlow *et al.*, 1983; Askenfelt and Hammarberg, 1986). However, only recently have these phenomena been interpreted as manifestations of nonlinear dynamics, i.e., bifurcations and chaos (Mende *et al.*, 1990; Herzel and Wendler, 1991; Titze *et al.*, 1993; Herzel *et al.*, 1994). A bifurcation consists of a jump from one oscillatory state to another, for a very small change in some system parameter such as muscle tension or subglottal pressure. The theory of nonlinear dynamics predicts that bifurcations are often accompanied by low-dimensional chaos. That is, the dynamics may appear to be unpredictable (stochastic), but they can nevertheless be explained deterministically by the nonlinear interactions of just a few variables. The acoustic signals of many types of rough voice have been analyzed using spectrograms, fundamental frequency contours, phase portraits, Poincaré sections, estimations of attractor dimensions, and estimations of Lyapunov exponents. Repeatedly, these studies have shown that many irregular voice signals are consistent with the output of a low-dimensional dynamical system (Herzel *et al.*, 1994; Narayanan and Alwan, 1995; Kumar and Mulik, 1996; Behrman and Baken, 1997; Behrman, 1999).

In addition to the bifurcations noted in rough voice signals, similar bifurcation scenarios have been observed in a variety of low-order models of vocal fold vibration (Ishizaka and Isshiki, 1976; Herzel *et al.*, 1991; Wong *et al.*, 1991; Smith *et al.*, 1992). Even bifurcations noted in high-order finite element simulations of vocal fold vibration have been shown to be governed by underlying low-order dynamics (Berry *et al.*, 1994; Herzel *et al.*, 1995).

This evidence of low-dimensionality in irregular vocal fold vibration is significant. It suggests that many of the irregularities in rough, disordered voice may be conceptualized in terms of low-order, nonlinear laryngeal models. Significantly, low-dimensional dynamics facilitates the task of establishing basic mechanisms of irregular vocal fold vibration.

Perhaps the greatest deterrent to establishing basic mechanisms of irregular vocal fold oscillations is the difficulty in making quantitative observations of the vibration patterns. The small size and inaccessibility of the larynx limit

[a)]Author to whom correspondence should be addressed. Address for correspondence: UCLA School of Medicine, Division of Head and Neck Surgery, 31-24 Rehabilitation Center, 1000 Veteran Ave., Los Angeles, CA 90095-1794; electronic mail: daberry@ucla.edu

the possibilities for acquiring such observations. Consequently, many studies of vocal fold vibration have relied on the study of secondary signals, such as the acoustic signal, the electroglottogram (EGG), and the glottal volume velocity. These signals are termed ''secondary,'' because if vocal fold vibration is the focus of the study, the signals provide indirect, as opposed to direct, measures of tissue vibration. Although such signals contain information about the glottal source and may be collected noninvasively, it is not reasonable to suppose that they contain enough information to regenerate the complex vibration patterns of the vocal folds. Like most inverse problems in the physical sciences, the regeneration of vocal fold vibrations from secondary signals is an ill-defined task with multiple solutions. To solve the problem, one needs to make many assumptions about the source, which may or may not be valid.

Videostroboscopy and high-speed endoscopy provide more direct information about vocal fold vibration. They yield unique information about the static and vibrating folds that is not available from secondary signals. The pictures and movies of vocal fold vibration captured by endoscopy have helped clinicians gain a greater intuitive understanding of the dynamics of vibration and the etiology of disorders. However, even endoscopy has its limitations. The superior view captured by endoscopy conceals the medial surface of the vocal folds during glottal closure and precludes direct measurement of vertical vibration.

Because of this endoscopic orientation, when clinicians talk about the mucosal wave, they usually refer to its propagation along the superior surface of the vocal folds, which is where they see the wave most clearly. However, upon reaching the superior surface of the folds, the mucosal wave attenuates quickly as it travels laterally outward. In terms of its influence on voice production, the most critical region of mucosal wave propagation occurs along the medial surface of the vocal folds: the mucosal wave is initiated along the medial-inferior surface of the vocal folds; it travels a significant distance along this surface; and it exhibits its greatest amplitude excursions along this surface. While traveling up this medial surface, the mucosal wave interacts nonlinearly with the opposite vocal fold during collision. It also experiences possible nonlinear interactions with sub- and supraglottal systems as it modulates the glottal airflow, and produces the acoustic wave, which is later perceived as voice. If there were any region of the vocal folds that should be imaged to better understand mechanisms of vibration, propagation of the mucosal wave, and its influence on voice, it would be the medial surface of the vocal folds. Unfortunately, the medial surface of the vocal folds is imaged very poorly with endoscopy.

To date, very few quantitative studies have been performed of the medial surface dynamics of the vocal folds. Even for regular, quasi-periodic vibrations there are only a few notable exceptions, including the excised larynx studies by Baer (1975) and the pellet-tracking experiments by Saito *et al.* (1981, 1985). Recent attempts have also been made to acquire quantitative observations of vocal fold movement using electronic speckle pattern interferometry (Gardner *et al.*, 1992, 1995).

Using a microscope and stroboscopy, Baer (1975) was able to track fleshpoint trajectories of excised canine larynges during phonation. Fleshpoints were marked with small particles, each possessing a diameter of about 0.08 mm. Both superior and inferior views were used, allowing fleshpoints to be tracked throughout most of the glottal cycle. In order to permit an inferior view, laryngeal airflow was introduced laterally, and the trachea was sealed with transparent plexiglass.

Vertical motion was detected by focusing the lens of the microscope on the observation particles and recording the focal length. Medial movements were acquired by centering the cross-hairs of the microscope on the particles, and noting the position coordinates. ''Still motion'' was obtained by strobing at the vibration frequency of the folds, giving the observer opportunity to focus the lens and adjust the crosshairs of the microscope. A strobe advance allowed the user to step through the periodic vibrations at $\frac{1}{8}$-cycle increments. Finally, transcriptions were converted to physical coordinates by employing appropriate transformations. With this procedure, it usually took about 15 min of observation time to record one cycle of vibration (Baer, 1975).

As the first quantitative study of its kind, Baer's study was a landmark in vocal fold research. However, some of the limitations of the Baer (1975) study included the following.

(1) The maximum number of fleshpoints tracked simultaneously was 3; thus, the vibratory patterns of coronal cross-sections of the vocal folds could be only crudely postulated.
(2) Only periodic or quasi-periodic vibrations could be observed using this stroboscopic procedure, i.e., irregular oscillations could not be observed.
(3) Because of dehumidification, there were changes in the vibratory patterns of the folds over the 15 min taken to record a complete vibratory cycle. Fundamental frequency changes were noted as well. Thus, the recorded trajectories could only be taken as rough approximations of vocal fold movement, averaged over thousands of cycles with questionable stationarity.
(4) Using excised larynges, the influence of the TA (thyroarytenoid) muscle activity could not be incorporated.
(5) Because of histological differences between canine and human vocal folds, the vibratory patterns of canine and human folds might be expected to differ.

Saito *et al.* (1985) used x-ray stroboscopic techniques to view vocal fold vibrations from a coronal aspect. Although stroboscopic x-ray laminagraphy had been performed previously (Hollien *et al.*, 1968) the Saito *et al.* (1985) study was one of the first which was quantitative in nature. Both surface and interior vocal fold fleshpoints were tracked by placing lead pellets (approximately 0.5 mm in diameter) in the vocal fold tissues. One of the advantages of this study was that it could be performed on anesthetized canines *in vivo*. However, because the density of lead was approximately 11.3 times greater than the density of the folds, one was left to wonder how seriously the vocal fold oscillations were altered by the artificial mass redistribution. Also, because of the large pellet size, the resolution of the trajectories was

such that the results had more qualitative than quantitative significance.

While electronic speckle pattern interferometry (ESPI) has traditionally been a technique for quantifying small static deflections, Gardner *et al.* (1995) were able to image the vibrating larynx by employing a stroboscopic implementation of the technique. However, the stroboscopic implementation limited the technique to a study of periodic or quasiperiodic vibrations; irregular vibrations could not be imaged. Moreover, while it was theoretically possible to get quantitative information from the interference fringes, the strobed images of periodic vibrations were generally of poor quality, and the technique needed further refinement before it was ready for practical use (Gardner *et al.*, 1995).

A hemilarynx investigation by Jiang and Titze (1993) demonstrated the potential to enhance the observation and quantification of the dynamics of the medial surface of the vocal folds. This study demonstrated the feasibility of using a hemilarynx as a substitute for a full larynx in experimental situations. Across a wide range of subglottal pressures, the hemilarynx and the full larynx yielded comparable amplitudes of vibration and fundamental frequencies, i.e., they never differed by more than about 10%. In addition, the average airflow of the hemilarynx was approximately one-half the airflow of the full larynx, as expected. The hemilarynx setup was realized by removing the right fold of the larynx, and replacing it with a vertical plexiglass plate. Building upon this former study, the present investigation also employs a hemilarynx methodology, yielding a sagittal view of the folds. Using this technique, vocal fold fleshpoints on the medial surface of the folds can be viewed throughout the entire glottal cycle, even during glottal closure.

The present investigation also builds upon Baer's quantitative study of laryngeal vibration in excised larynges (Baer, 1975). New experimental methods and technological advances were used to perform quantitative observations of both regular and irregular vocal fold oscillations. Specifically, the limitations of the Baer (1975) study were addressed as follows:

(1) The number of fleshpoints tracked simultaneously was increased from three to nine. The increased spatial resolution permitted more detailed reconstructions of the vibratory patterns of coronal cross-sections.
(2) High-speed digital imaging was used. With increased temporal resolution, aperiodic oscillations could be imaged. Thus, mechanisms of irregular oscillation could be probed.
(3) Dehumidification artifacts were virtually eliminated since the recording procedure was almost instantaneous. Also, because strobing was not used, there was no danger of averaging over changing vibration patterns.
(4) In the excised larynx investigations, the influence of the TA muscle activity was still neglected. Although phonation involves many neurologic and biomechanical processes, this investigation focused exclusively on the physical plant, the biomechanics of tissue and air in motion. The dynamics of laryngeal tissues in an excised larynx experiment, like the corresponding tissues in an intact physical system, are governed by the basic laws of flow-induced oscillation. Even with only passive tissues properties present, many aspects of the intact voice are routinely duplicated in excised larynx experiments, including chestlike vibrations, falsettolike vibrations, frylike vibrations, and many of the bifurcations and irregularities often associated with voice disorders (Baer, 1975; Jiang and Titze, 1994; Berry *et al.*, 1996). Furthermore, the excised larynx setup, and in particular the hemilarynx setup, provided additional access for viewing and quantification of laryngeal vibration that would otherwise not be possible.
(5) Although excised canine larynges were the focus of this study, the methods could also be extended to studies of excised human larynxes. The canine larynx has been studied for years as a model comparable to the human larynx. The results of this study provided significant data for comparison with previous investigations of excised canine larynxes.

In this report, we present the methodology for high-speed, quantitative imaging of the medial surface of vocal fold tissues in excised larynx experiments, along one coronal cross-section of the left vocal fold. Quantitative results will be reported for a single experiment, and empirical eigenfunctions will be extracted. Empirical eigenfunctions will be given a precise, mathematical definition in Sec. III. Conceptually, they may be viewed as the basic building blocks of many simple and complex vibration patterns. In previous computational studies, just a few empirical eigenfunctions generated a variety of periodic and aperiodic vocal fold vibrations (Berry *et al.*, 1994). In this laboratory investigation, the medial surface vibrations of vocal fold tissues were imaged with sufficient temporal and spatial resolution to enable extraction of empirical eigenfunctions. The eigenfunctions from the excised larynx experiment will be compared with those reported in previous computational studies (Berry *et al.*, 1994; Alipour *et al.*, 2000). Quantitative investigations of this type continue to provide crucial information for exploring basic mechanisms of regular and irregular vocal fold vibration (Berry *et al.*, 1994; Neubauer, *et al.*, 2001).

II. METHODS

A canine larynx was obtained from an experimental animal weighing approximately 23 kg. The animal was sacrificed for other purposes, but the tissue was made available to us post mortem. Immediately after harvest, the larynx was refrigerated and stored in saline solution. The experiment was conducted 5 days after harvest. The larynx was dissected to remove the epiglottis, the ventricular folds, and all but a short section of the trachea (approximately 2 in.). In order to create a hemilarynx, a vertical slice was made along the glottal midline. The depth of the slice was approximately 15 mm below the superior surface of the folds. The right fold and portions of the surrounding cricoid and thyroid cartilages were removed. So that specific vocal fold fleshpoints could be tracked during the experiments, small markers were placed on the left vocal fold. In particular, 9-0 black monofilament nylon microsutures were placed along the me-

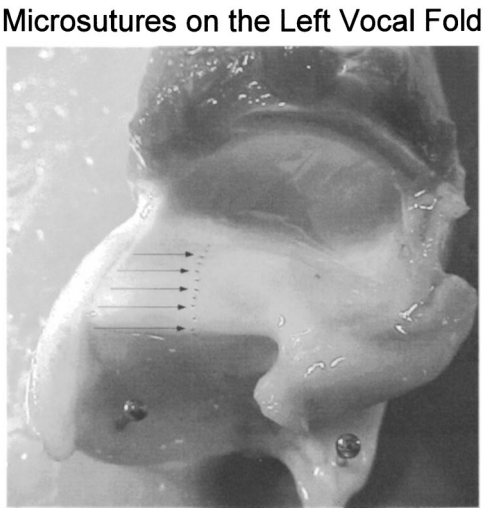


FIG. 1. As indicated by the arrows, individual fleshpoints were marked along the medial surface of one coronal cross-section of the left vocal fold using microsutures. Approximately a 1-mm spacing exists between sutures.

dial surface of a mid-coronal section of the left fold, as shown in Fig. 1. The microsutures had an average diameter of approximately 0.034 mm (Ethicon, Inc.), and the spacing between microsutures was approximately 1 mm. The microsutures were positioned by an experienced phonosurgeon who attempted to penetrate just the mucosal epithelium [which has a thickness of approximately 0.05–0.10 mm (Hirano, 1977)], while avoiding penetration of the lamina propria. Visual comparisons of tissue vibrations in regions with and without microsutures revealed no alterations in the vibration patterns.

The left fold was mounted against a glass plate, as shown in Fig. 2(a). A silicone sealant was used to seal any airflow leaks between the tissue, cartilage, and glass. A prism was placed on the other side of the glass plate, yielding two oblique views of the medial surface of the left vocal fold. Without the oblique views, lateral vibrations could not have been detected. The two faces of the prism had a height and width of 1 in., and the angle between the two faces was 90 degrees. Finally, a Kodak EktaPro 4540 high-speed digital imaging system imaged the vibrations through the two faces

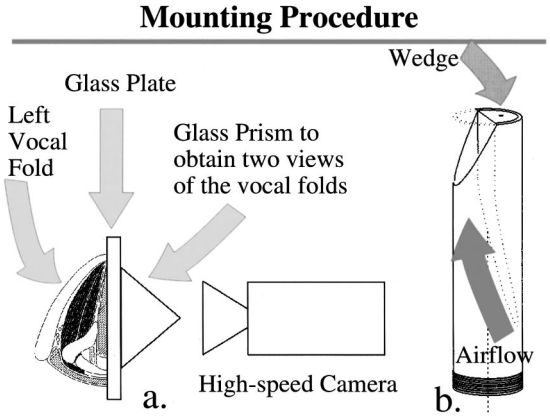


FIG. 2. (a) A superior view of the hemilarynx setup, including the left vocal fold, the glass plate, the glass prism, and the high-speed camera. (b) The artificial trachea upon which the hemilarynx is mounted for phonation.

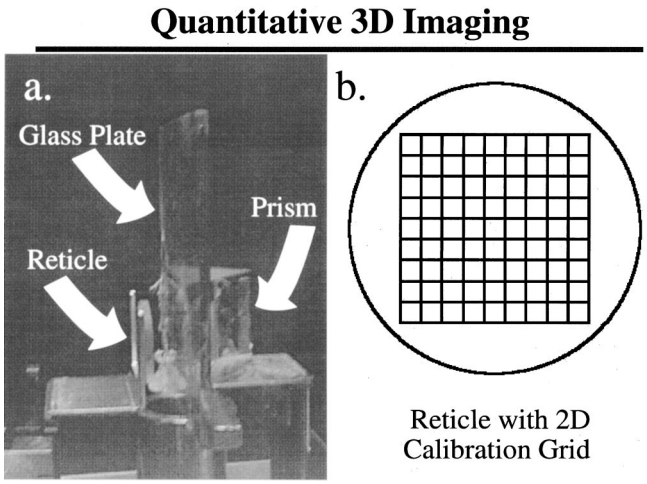


FIG. 3. (a) Illustration of the calibration technique used to obtain 3D coordinates from the images. (b) A view of the 2D grid shown on the reticle, which has a 1.27-mm spacing between horizontal and vertical lines.

of the prism. Vibrations were imaged at a sampling frequency of 4500 frames per second, with a total image resolution of 256×256 pixels. The image was black and white with 256 levels of grayscale. A 300-W bulb was placed approximately 1.5 ft above the larynx to provide additional illumination during the imaging process.

The larynx was also mounted on a stainless steel cylindrical tube, as shown in Fig. 2(b). A stainless steel wedge was also firmly mounted in the tube to smoothly channel the airflow beneath the left fold. Vibrations were induced by passing airflow up through the trachea and through the hemiglottis (i.e., the area between the glass plate and the left vocal fold).

Subglottal pressure was varied between 0 and 16 cm H_2O. Between these extremes, a variety of vibratory modes were imaged, including both periodic and aperiodic regimes, and transitions between these regimes. Among the periodic regimes, a falsettolike vibration was imaged at a subglottal pressure of approximately 6 cm H_2O, and chestlike vibration pattern at approximately 8 cm H_2O (Berry and Montequin, 1998). High-intensity, large-amplitude vibrations, including both periodic and aperiodic segments, were imaged at subglottal pressures ranging from 14 to 16 cm H_2O.

Following image capture, a 3D calibration grid was established so that image coordinates could be mapped onto physical coordinates. This was done by removing the hemilarynx from the glass plate, and replacing it with a glass reticle containing a 2D calibration grid. The positioning of the reticle against the glass plate is shown in Fig. 3(a), and a schematic of the reticle is shown in Fig. 3(b). The physical spacing between both horizontal and vertical lines on the reticle was 1.27 mm. The reticle was mounted against the glass plate using a micrometer. The reticle was placed flush against the glass plate and imaged. Using the micrometer, the reticle was perpendicularly displaced from the plate by 1.0 mm, and imaged again. This procedure was repeated up to a 10 mm displacement. In this way, an entire 3D grid was imaged through the two faces of the prism. The other imaging apparatus (glass plate, prism, camera) remained firmly

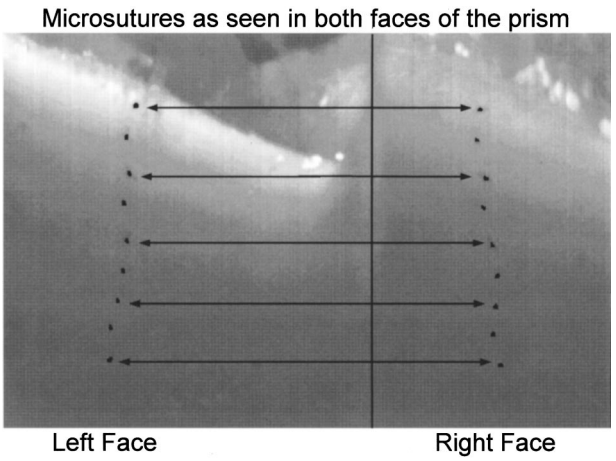


FIG. 4. The microsutures on the medial surface of one coronal cross-section of the left vocal, as imaged through the two faces of the glass prism. Arrows indicate the location of the microsutures in both views.

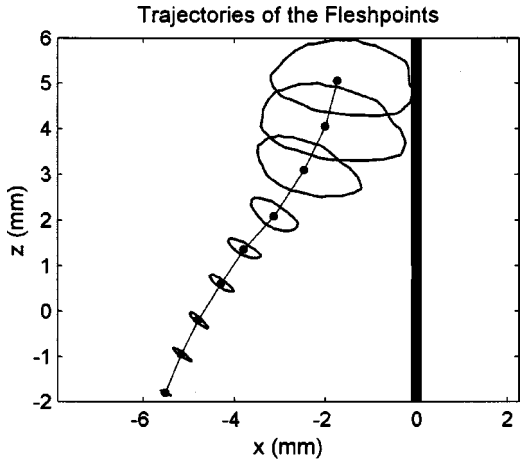


FIG. 5. An anterior view of the left vocal fold. The ellipses indicate the trajectories of vocal fold fleshpoints over one cycle of vibration, with physical dimensions labeled. All the fleshpoints travel in counter-clockwise orbits. The dark vertical line indicates the glass plate in the hemilarynx set up. The bolded dots, connected by a solid line, indicate the dynamic equilibria of the fleshpoints.

mounted, and undisturbed during this procedure, so as not to alter the calibration.

Using the direct linear transform (DLT), a precise mapping was established from image coordinate to physical coordinates (Abdel-Aziz and Karara, 1971; Shapiro, 1978). In particular, for the first oblique view, the image coordinates x_i and y_i were mapped onto the physical coordinates X_i, Y_i, and Z_i using the following two relations:

$$\begin{aligned} &x_i + L_1 X_i + L_2 Y_i + L_3 Z_i + L_4 + L_9 x_i X_i + L_{10} x_i Y_i \\ &\quad + L_{11} x_i Z_i = 0, \\ &y_i + L_5 X_i + L_6 Y_i + L_7 Z_i + L_8 + L_9 y_i X_i + L_{10} y_i Y_i \\ &\quad + L_{11} y_i Z_i = 0. \end{aligned} \tag{1}$$

The L_i are constants related to camera orientation, prism orientation, and lens distortion. To compute the physical coordinates, a corresponding set of relations was also applied for the second oblique view.

With 11 unknown constants per view, at least 6 known grid points were required to compute a DLT solution. As already described, a procedure using a reticle supplied known physical data points with which to compute the transform coefficients L_i. Corresponding image coordinates were obtained by displaying the reticle images on a computer monitor, and clicking a mouse at the intersections of the grid lines of the reticle. The standard deviation in making repeated measurements of the intersections of the grid lines was 0.30 pixels in image coordinates. With known physical coordinates and computed image coordinates, the transformation coefficients L_i were determined using a least square error procedure. Across the grid points considered, the root-mean-square (rms) error of the transform was 0.081 mm.

A sample image captured by the high-speed imaging system is shown in Fig. 4. Only one vocal fold was imaged, but the glass prism yielded two oblique views of the left vocal fold. The microsutures may be seen in both faces of the prism. Indeed, to compute the physical coordinates of a microsuture, the microsuture location had to be identified in both faces of the prism so that image coordinates could be specified. With the image coordinates, and the previously determined transformation constants L_i, the physical coordinates of the microsutures were determined using Eq. (1).

Image coordinates were obtained using a semi-automatic marker tracking program. Initially, the computer operator would click on a microsuture to track in either the left or right face of the prism. The computer program would locate the lowest intensity pixel within ± 3 pixels of the mouse click. For each subsequent image, the computer program would estimate the subsequent suture location by finding the lowest intensity pixel in the new image within ± 3 pixels of the previous pixel location. The computer operator could either accept this value, or click on another value, which the computer program would adjust based on the lowest intensity pixel within ± 3 pixels of the mouse click. Because the computer program determined pixel location based on a lowest intensity criteria, the results of the marker tracking procedure were repeated with identical results. However, spatial resolution was limited by the pixel resolution, yielding spatial quantization error. In subsequent studies, spatial interpolation could be used to increase the spatial resolution. After trajectories of all the sutures were obtained from both left and right views, image coordinates were converted into physical coordinates using Eq. (1).

III. RESULTS AND DISCUSSION

Figure 5 shows an anterior view of the elliptical paths traversed by the fleshpoints of the left vocal fold over one cycle of vibration. This was a periodic vibration pattern with a fundamental frequency of approximately 102 Hz, generated with a subglottal pressure of 8 cm H_2O. The amplitudes of vibration of the microsutures from superior to inferior positions were roughly 3.2, 3.2, 2.4, 1.1, 0.77, 0.57, 0.51, 0.46, and 0.23 mm, as measured along the major axis of each ellipse. Because the suture trajectories are displayed in a coronal plane, only the medio-lateral (x) and inferior-superior (z) vibrations are illustrated, not anterior-posterior (y) vibrations. However, similar to previous studies (Baer,

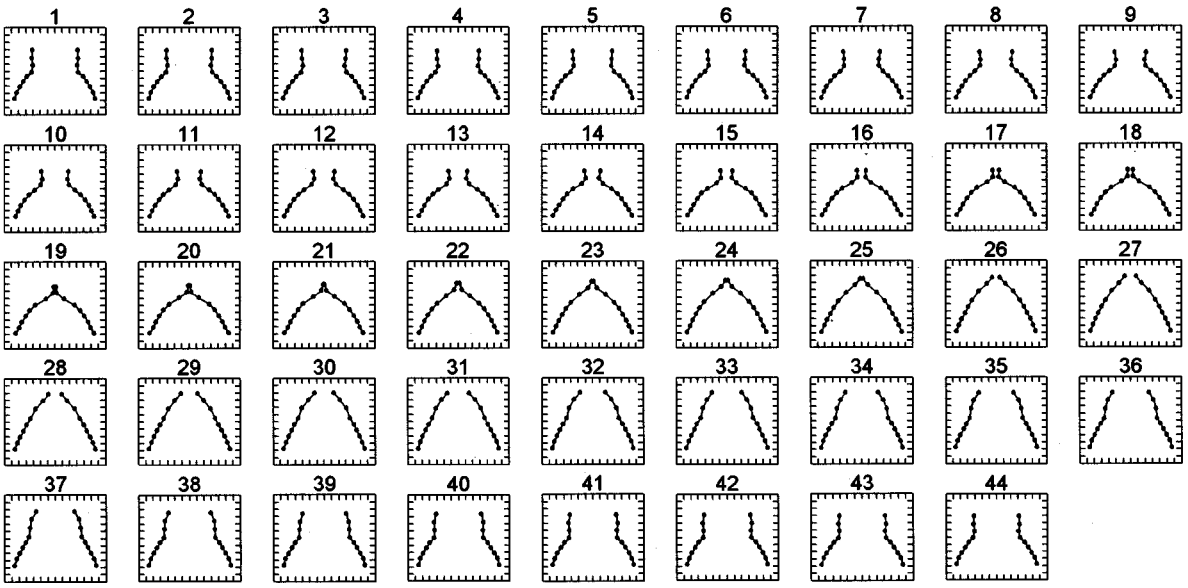


FIG. 6. From an anterior view, frames (1)–(44) show the physical coordinates of the vocal fold fleshpoints sequentially in time over one cycle of vibration. The fleshpoints are indicated by bolded dots, connected by solid lines. Although only the left vocal fold was imaged, the right vocal fold is also shown to illustrate the corresponding glottal shape on a symmetric full larynx. Ticks on the coordinate axes indicate a 1-mm spacing.

1975), this investigation confirms that the vibrations in the anterior-posterior direction are about an order of magnitude smaller than the vibrations in the other two directions. Figure 6 depicts the same data in successive temporal frames, as tracked over one glottal cycle, or roughly 44 frames. From an anterior view, this figure illustrates the progression of the mucosal wave as it travels up the medial surface of the vocal folds. The vibrations illustrate a sharp, whiplike closure of the vocal fold, reminiscent of the chest or modal register (Berry and Montequin, 1998).

Using the quantitative data from this vibration pattern, the method of empirical eigenfunctions was employed to examine basic mechanisms of vibration. First, the physical coordinates R_i of the fleshpoints were separated into mean and time-varying components:

$$R_i(t_k)=\overline{R_i}+r_i(t_k), \tag{2}$$

where the bar denoted a mean value, t_k represented the discrete times at which the coordinates were sampled, and i was a dummy index to distinguish the vocal fold fleshpoints. The physical coordinates R_i included the set of X_i (lateral coordinates) and Z_i (vertical coordinates). Longitudinal coordinates Y_i were neglected to facilitate comparison with previous results from computer models which explicitly assumed that no longitudinal displacements were present in the vibration patterns. Also, as already indicated in this study, longitudinal displacements were an order of magnitude smaller than displacements in the other two directions.

The components S_{ij} of the covariance matrix S were generated using the time-varying coordinates r_i:

$$S_{ij}=\frac{1}{N}\sum_{k=1}^{N} r_i(t_k)r_j(t_k), \tag{3}$$

where N represented the total number of time samples. The eigenvectors of the covariance matrix yielded the empirical eigenfunctions of the system, and the eigenvalues explained the relative amount of energy captured by each eigenfunction.

For the vibrations under consideration, two eigenfunctions explained 98% of the energy of the vocal fold vibrations. A comparison of the original vibrations with a superposition of the first two eigenfunctions yielded a rms error of 0.084 mm, suggesting that the remaining 2% of the energy was attributable to random errors in measurement, and/or the coordinate transformation. The first eigenfunction captured 69% of the energy, and the second eigenfunction explained 29%. The two eigenfunctions are shown in Fig. 7. This is similar to the Berry *et al.* (1994) theoretical study, in which

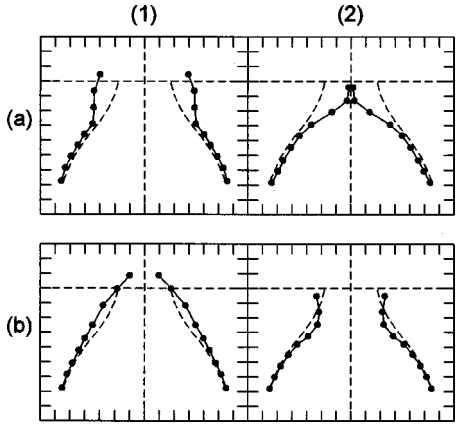


FIG. 7. An anterior view of the eigenfunction shapes, as indicated by bolded dots, and connected by solid lines. Columns (1) and (2) illustrate the eigenfunctions at maximum and minimum excursions. The eigenfunctions may be viewed as excursions about the dynamic equilibrium of the vocal fold surface, which is indicated as a curved dashed line. (a) The first eigenfunction captures the lateral vibrations of the folds and explains 69% of the variance. (b) The second eigenfunction captures glottal shaping between convergent and divergent states, and explains 29% of the variance. The vertical dashed line indicates the glottal midline and the horizontal dashed line indicates the top of the glottis (i.e., the vertical dynamic equilibrium of the top microsuture). Ticks on the coordinate axes indicate a 1-mm spacing.

the two dominant eigenfunctions from a finite element model captured 73% and 25% of the energy, respectively, for a total of 98%.

In the current study, the first eigenfunction captured the net lateral vibrations of the vocal folds, as shown in Fig. 7(a). This figure illustrates the eigenfunction at opposite extremes in the glottal cycle, one at maximum opening [Fig. 7(a), column 1], and one at minimum opening [Fig. 7(a), column 2]. In comparison to the two-mass model (Ishizaka and Flanagan, 1972), this eigenfunction is related to the mode in which both top and bottom masses vibrate in-phase with each other. It is the mode responsible for modulating the glottal airway, and generating the acoustic signal.

The second eigenfunction captured the glottal shaping of the vocal folds, alternately shaping a convergent [Fig. 7(b), column 1] and divergent glottis [Fig. 7(b), column 2]. This is especially true for the top two or three fleshpoints. A convergent glottis is one in which the superior glottal space is smaller than the inferior glottal space. That is, as one traverses the glottal airway from bottom to top, the glottal airway narrows, or converges. A divergent glottis is just the opposite, meaning that the glottal airway increases as one traverses the glottal airway from bottom to top. It is known that a divergent glottis possesses a relatively low intraglottal pressure and that a convergent glottis possesses a relatively high intraglottal pressure (Titze, 1988). Thus, this eigenfunction is intimately related to the intraglottal pressure.

In addition, notice that the second eigenfunction was a predominantly a rotational mode of vibration, capturing roughly equal amounts of lateral and vertical motion. For example, a convergent glottal shape was associated with an elevated vertical position, and a divergent shape was associated with a lowered vertical position. This phenomenon was also observed in the finite element model (Berry *et al.*, 1994). Physically, it makes sense that the vocal folds should rise during a convergent glottal state. Because the intraglottal pressure pushes on the vocal folds in a direction perpendicular to their medial surface, as the vocal folds begin to converge, the intraglottal pressure pushes the vocal folds not only outward, but also upward.

This investigation suggests that the *singular* role of the first eigenfunction is to execute the primary function of vocal fold vibration: the generation of the acoustic signal through modulation of the glottal airflow. In this example, the first eigenfunction captured all the lateral vibrations in which superior and inferior margins of the folds were predominantly in-phase with one another. All other movements of the vocal folds were considered secondary, and thus captured by the second eigenfunction. In this case, the second eigenfunction captured glottal shaping (convergence and divergence of the folds), as well as the vertical oscillations associated with glottal shaping. In other words, the second eigenfunction played a supportive role, e.g., it did not modulate the glottal airflow, but it produced favorable physical conditions so that energy transfer could occur from the airflow to the tissues. Favorable physical conditions were produced through appropriate glottal shaping.

To facilitate self-oscillation of the vocal folds, the energy transfer from the glottal airflow to the vocal fold tissues needs to be maximized (Stevens, 1977; Broad, 1979; Titze, 1988; Berry *et al.*, 1994). This conditions occurs when the intraglottal pressure (in this case, controlled by eigenfunction 2) is in-phase with the net lateral velocity of the folds (in this case, controlled by eigenfunction 1). This phenomenon is illustrated in Fig. 8. For example, as the vocal folds approach the glottal midline [Fig. 8(b), frames 1–4], a low-pressure condition is created by a divergent glottis [Fig. 8(c), frames 1–4], thus ''helping'' the folds to approximate. On the other hand, as the vocal folds separate [Fig. 8(b), frames 5–8], a high-pressure condition is created by a convergent glottis [Fig. 8(c), frames 5–8], thus ''helping'' to push the folds apart. Thus, a specific entrainment or temporal phase relationship between the eigenfunctions facilitated self-oscillation. Such relationships have already been discussed in relationship to eigenfunctions extracted from a finite ele-

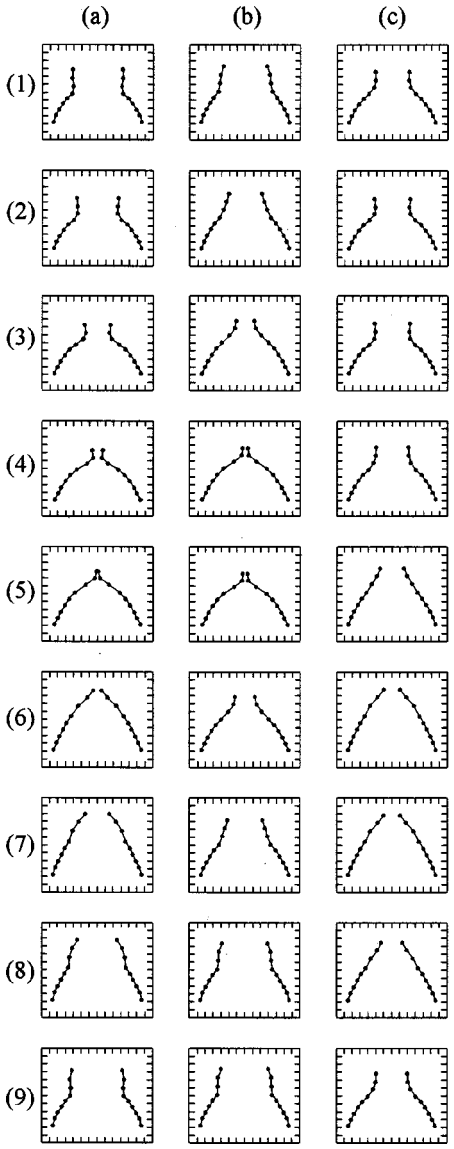


FIG. 8. The sequential time frames (1)–(9) illustrate the temporal phase relationship between the (a) original vocal fold vibrations, (b) eigenmode 1 and, and (c) eigenmode 2, over one cycle of vibration, from an anterior view. Because of space limitations, only the data from every fifth temporal frame of the high-speed movie is shown. Ticks on the coordinate axes indicate a 1-mm spacing (after Berry, in press).

ment simulation of vocal fold movement (Berry *et al.*, 1994). However, such hypotheses concerning self-oscillation have never been studied previously on actual vibrating tissues.

While conceptually similar, the eigenfunction analysis from this investigation was not identical to that computed from the finite element investigation. Although the two eigenfunctions captured 98% of the energy for both theory and experiment, the primary and secondary eigenfunctions were switched in the finite element model, as compared to the current empirical investigation. One reason for this may be that the tissue simulated in Berry *et al.* (1994) was highly compressible, a condition that was rectified in subsequent modeling studies (Berry and Titze, 1996; Alipour *et al.*, 2000).

IV. SUMMARY

In conclusion, a method for high-speed quantitative imaging was demonstrated on an excised canine larynx in the laboratory. Nine individual fleshpoints were tracked on the medial surface of the left vocal fold in a coronal plane, approximately midway between anterior and posterior extremes, with approximately a 1-mm spacing between marked fleshpoints. Frames were imaged at a rate of 4500 frames/s using a Kodak EktaPro 4540 high-speed digital imaging system. Individual fleshpoints were tracked with a spatial resolution of about 0.08 mm. With this spatio-temporal resolution, empirical eigenfunctions were extracted from vocal fold tissues, illuminating the underlying building blocks of the vibration patterns. The analysis indicated that a typical case of chestlike vibration patterns of the vocal folds could be explained by just two dominant eigenfunctions. Though the eigenfunctions were not identical with theoretical predictions, there were many qualitative similarities between previous analytic (Berry and Titze, 1996; Berry, in press) and computational studies (Berry *et al.*, 1994; Alipour *et al.*, 2000). Differences between the two may suggest limitations in current vocal fold models.

Previous theoretical studies (e.g., Berry *et al.*, 1994; Herzel *et al.*, 1995) have indicated that even many complicated vibration patterns can be explained by just a few underlying eigenfunctions. In this study, while many aperiodic vibration patterns have been imaged, the marker tracking procedure for the aperiodic sequences has not yet been completed. For such aperiodic sequences, literally thousands of images need to be analyzed. An automatic marker tracking routine, as opposed to a semi-automatic routine, would be required to make such a task realizable.

Nevertheless, the imaging procedure outlined in this investigation yields high-speed quantitative data which may be used to help probe mechanisms of regular and irregular vocal fold vibration. As stated earlier, perhaps the greatest deterrent to establishing mechanisms of regular and irregular vocal fold oscillations has been the difficulty in making quantitative observations of the medial surface of the vocal folds. The purpose of the present investigation has been to help remedy this situation.

ACKNOWLEDGMENTS

This work was supported by Grant No. R29 DC03072 from the National Institute on Deafness and Other Communication Disorders. The authors also thank Brad H. Story and Martyn J. Clark who made important contributions to the experimental design used in this investigation.

Abdel-Aziz, Y. I., and Karara, H. M. (**1971**). ''Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry,'' *in Proceedings of the ASP/UI Symposium on Close-Range Photogrammetry* (American Society of Photogrammetry, Falls Church, VA), pp. 1–18.

Alipour, F., Berry, D. A., and Titze, I. R. (**2000**). ''A finite element model of vocal fold vibration,'' J. Acoust. Soc. Am. **108**, 3003–3012.

Askenfelt, A. G., and Hammarberg, B. (**1986**). ''Speech waveform perturbation analysis: A perceptual-acoustical comparison of seven measures,'' J. Speech Hear. Res. **29**, 50–64.

Baer, T. (**1975**). ''Investigation of phonation using excised larynges,'' Ph.D. dissertation, Massachusetts Institute of Technology, Boston, MA.

Behrman, A. (**1999**). ''Global and local dimensions of vocal dynamics,'' J. Acoust. Soc. Am. **105**, 432–443.

Behrman, A., and Baken, R. J. (**1997**). ''Correlation dimension of electroglottographic data from healthy and pathologic subjects,'' J. Acoust. Soc. Am. **102**, 2371–2379.

Berry, D. A. (**in press**). Mechanisms of modal and nonmodal phonation, J. Phonetics.

Berry, D. A., and Montequin, D. W. (**1998**). ''Contrasting chest and falsetto-like vibration patterns of the vocal folds,'' in *Proceedings of the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America*, Vol. 4, pp. 2667–2668.

Berry, D. A., and Titze, I. R. (**1996**). ''Normal modes in a continuum model of vocal fold tissues,'' J. Acoust. Soc. Am. **100**, 3345–3354.

Berry, D. A., Herzel, H., Titze, I. R., and Krischer, K. (**1994**). ''Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions,'' J. Acoust. Soc. Am. **95**, 3595–3604.

Berry, D. A., Herzel, H., Titze, I. R., and Story, B. H. (**1996**). ''Bifurcations in excised larynx experiments,'' J. Voice **10**, 129–138.

Broad, D. (**1979**). ''The new theories of vocal fold vibration,'' in *Speech and Language: Advances in Basic Research and Practice*, edited by N. Lass (Academic, New York).

Gardner, G. M., Castracane, J., Conerty, M., and Parnes, S. M. (**1992**). ''Laser holography and interferometry of the larynx: preliminary studies with a cadaver model,'' Trans. Am. Laryngol. Assoc. **113**, 104–109.

Gardner, G. M., Castracane, J., Conerty, M., and Parnes, S. M. (**1995**). ''Electronic speckle pattern interferometry of the vibrating larynx,'' Ann. Otol. Rhinol. Laryngol. **104**, 5–12.

Herzel, H., and Wendler, J. (**1991**). ''Evidence of chaos in phonatory samples,'' in Proceedings of EUROSPEECH (European Speech Communication Association, Genova, Italy), pp. 263–266.

Herzel, H., Berry, D. A., Titze, I. R., and Saleh, M. (**1994**). ''Analysis of vocal disorders with methods from nonlinear dynamics,'' J. Speech Hear. Res. **37**, 1008–1119.

Herzel, H., Krischer, K., Berry, D. A., and Titze, I. R. (**1995**). ''Analysis of spatio-temporal patterns by means of empirical orthogonal eigenfunctions,'' in *Spatio-temporal Patterns in Nonequilibrum Complex Systems*, edited by P. E. Cladis and P. Palffy-Mulhoray (Addison–Wesley, Reading, MA), pp. 505–518.

Herzel, H., Steinecke, I., Mende, W., and Wermke, K. (**1991**). ''Chaos and bifurcations during voiced speech,'' in *Complexity, Chaos, and Biological Evolution*, edited by E. Mosekilde (Plenum, New York), pp. 41–50.

Hirano, M. (**1997**). ''Structure and vibratory behavior of the vocal folds,'' in *Dynamic Aspects of Speech Production*, edited by M. Sawashima and S. C. Franklin (Univ. of Tokyo, Tokyo), pp. 13–30.

Hollien, H., Coleman, R., and Moore, P. (**1968**). ''Stroboscopic laminagraphy of the larynx during phonation,'' Acta Otolaryngol. **65**, 209–215.

Ishizaka, K., and Flanagan, J. L. (**1972**). ''Synthesis of voiced sounds from a two-mass model of the vocal cords,'' Bell Syst. Tech. J. **51**, 1233–1267.

Ishizaka, K., and Isshiki, N. (**1976**). ''Computer simulation of pathological vocal-cord vibration,'' J. Acoust. Soc. Am. **60**, 1193–1198.

Jiang, J. J., and Titze, I. R. (**1993**). ''A methodological study of hemilaryngeal phonation,'' Laryngoscope **103**, 872–882.

Jiang, J. J., and Titze, I. R. (**1994**). ''Measurement of vocal fold intraglottal pressure and impact stress,'' J. Voice **8**, 132–144.
Koike, Y. (**1969**). ''Amplitude modulations in patients with laryngeal diseases,'' J. Acoust. Soc. Am. **45**, 839–844.
Kumar, A., and Mullick, S. K. (**1996**). ''Nonlinear dynamical analysis of speech,'' J. Acoust. Soc. Am. **100**, 615–629.
Ludlow, C., Couter, D., and Gentages, F. (**1983**). ''The differential sensitivity of frequency perturbation to laryngeal neoplasms and neuropathologies,'' in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. Bless and J. Abbs (College Hill, San Diego, CA), pp. 381–392.
Mende, W., Herzel, H., and Wermke, K. (**1990**). ''Bifurcations and chaos in newborn cries,'' Phys. Lett. A **145**, 418–424.
Narayanan, S. S., and Alwan, A. A. (**1995**). ''A nonlinear dynamical systems analysis of fricative consonants,'' J. Acoust. Soc. Am. **97**, 2511–2524.
Neubauer, J., Mergell, P., Eysholdt, U., and Herzel, H. (**2001**). Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial modes, J. Acoust. Soc. Am. (to be published).
Saito, S., Fukuda, H., Isogai, Y., and Ono, H. (**1981**). X-ray stroboscopy, in *Vocal Fold Physiology*, edited by K. Stevens (Univ. of Tokyo, Tokyo), pp. 95–106.
Saito, S., Fukuda, H., Kitahira, S., Isogai, Y., Tsuzuki, T., Muta, H., Takyama, E., Fujika, T., Kokawa, N., and Makino, K. (**1985**). ''Pellet tracking in the vocal fold while phonating: experimental study using canine larynges with muscle activity,'' in *Vocal Fold Physiology*, edited by I. R. Titze and R. C. Scherer (Denver Center for the Performing Arts, Denver, CO), pp. 169–182.
Shapiro, R. (**1978**). ''Direct linear transformation method for three-dimensional cinematography,'' Res. Q. **49**, 197–205.
Sirviö, P., and Michelsson, K. (**1976**). ''Sound-spectrographic cry analysis of normal and abnormal newborn infants,'' Folia Phoniatr. **28**, 161–173.
Smith, M. E., Berke, G. S., Gerratt, B. R., and Kreiman, J. (**1992**). ''Laryngeal paralyses: Theoretical considerations and effects on laryngeal vibration,'' J. Speech Hear. Res. **35**, 545–554.
Stevens, K. N. (**1977**). ''Physics of laryngeal behavior and larynx modes,'' Phonetica **34**, 264–279.
Titze, I. R. (**1988**). ''The physics of small-amplitude oscillation of the vocal folds,'' J. Acoust. Soc. Am. **83**, 1536–1552.
Titze, I. R., Baken, R., and Herzel, H. (**1993**). ''Evidence of chaos in vocal fold vibration,'' in *Vocal Fold Physiology: Frontiers in Basic Science*, edited by I. R. Titze (Singular, San Diego, CA), pp. 143–188.
Wieser, M. (**1981**). ''Periodendaueranalyse bei spastischen dysphonien,'' Folia Phoniatr. **33**, 314.
Wong, D., Ito, M. R., Cox, N. B., and Titze, I. R. (**1991**). ''Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases,'' J. Acoust. Soc. Am. **89**, 383–394.

Aero-acoustics of silicone rubber lip reeds for alternative voice production in laryngectomees

Marein van der Torn, Hans F. Mahieu, and Joost M. Festen[a)]
Department of Otolaryngology, Vrije Universiteit Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands

(Received 23 January 2001; accepted for publication 26 June 2001)

To improve voice quality after laryngectomy, a small pneumatic sound source to be incorporated in a regular tracheoesophageal shunt valve was designed. This artificial voice source consists of a single floppy lip reed, which performs self-sustaining flutter-type oscillations driven by the expired pulmonary air that flows through the tracheoesophageal shunt valve along the outward-striking lip reed. In this *in vitro* study, aero-acoustic data and detailed high-speed digital image sequences of lip reed behavior are obtained for 10 lip configurations. The high-speed visualizations provide a more explicit understanding and reveal details of lip reed behavior, such as the onset of vibration, beating of the lip against the walls of its housing, and chaotic behavior at high volume flow. We discuss several aspects of lip reed behavior in general and implications for its application as an artificial voice source. For pressures above the sounding threshold, volume flow, fundamental frequency and sound pressure level generated by the floppy lip reed are almost linear functions of the driving force, static pressure difference across the lip. Observed irregularities in these relations are mainly caused by transitions from one type of beating behavior of the lip against the walls of its housing to another. This beating explains the wide range and the driving force dependence of fundamental frequency, and seems to have a strong effect on the spectral content. The thickness of the lip base is linearly related to the fundamental frequency of lip reed oscillation.
[DOI: 10.1121/1.1398053]

PACS numbers: 43.70.Aj, 43.75.Fg [AL]

I. INTRODUCTION

Despite many efforts towards organ preservation in the treatment of locally advanced cancer involving the larynx, 20 to 40 percent of all laryngeal cancer cases eventually come to total laryngectomy, i.e., removal of the entire larynx [Figs. 1(A)–(B)]. The chances of being cured from this malignant disease are relatively high. Consequently, many patients live a considerable number of years with the impairments associated with the laryngectomy, such as the loss of voice, the tracheostoma, and the loss of smell. Most problems of laryngectomized patients are related to the loss of voice. It is generally acknowledged that rapid and effective voice restoration is critical to the successful reduction of psychological, social, and economic setbacks due to postlaryngectomy aphonia.[1]

Over the past two decades, surgical voice restoration by means of a tracheoesophageal (TE) fistula incorporating a silicone rubber shunt valve[2] has become the standard approach to postlaryngectomy voice rehabilitation in most European institutions. By closing the tracheostoma with a finger, expiratory air is directed through this one-way valve from the trachea into the upper digestive tract, while preventing leakage of food and saliva from the pharynx into the airway [Fig. 1(C)]. The term ''voice prosthesis'' is widely used in medical literature when referring to TE shunt valves, although these devices do not actually produce sound. The airflow sets the closely approximated mucosal surfaces of the former upper esophageal sphincter, also called pharyngoesophageal (PE) segment, into vibration. This low-frequency sound can be used as substitute voice. Compared to other methods of alaryngeal voice production, such as electrolarynx (Servox®) and esophageal voice, a TE fistula enables louder phonation and better intelligibility;[3] additionally, it enables usually quick and trouble-free voice acquisition, higher speech rate, and more sustained phrasing.[4,5]

If, however, the tonus of the PE segment is too low to attain sufficient mucosal approximation, the resulting TE shunt valve voice will be weak and breathy, or merely a coarse whisper.[6,7] With sufficient approximation, the vibrating mass is often fairly large, which yields a low fundamental frequency (f_0). Female laryngectomees in particular often have severe problems accepting their low-pitched alaryngeal voice. The mean speaking f_0 of laryngeal female voices is 211 Hz (S.D.=23 Hz),[8] which decreases after laryngectomy and current voice rehabilitation to an unnaturally low mean speaking f_0 of 108 Hz (S.D.=28 Hz).[9]

In order to improve voice quality for these two groups of laryngectomees (females and those with a hypotonic PE segment), a small pneumatic sound source to be incorporated in a regular TE shunt valve was designed in cooperation with the Department of Biomedical Engineering at the University of Groningen.[10,11] To obtain a periodically changing flow resulting in air pressure variations, several sound-producing principles were taken into consideration for the artificial voice source: an air-reed, a siren, an inward-striking reed, and a lip reed. The major disadvantage of the air-reed prin-

[a)]Author to whom correspondence should be addressed; electronic mail: jm.festen@vmuc.nl

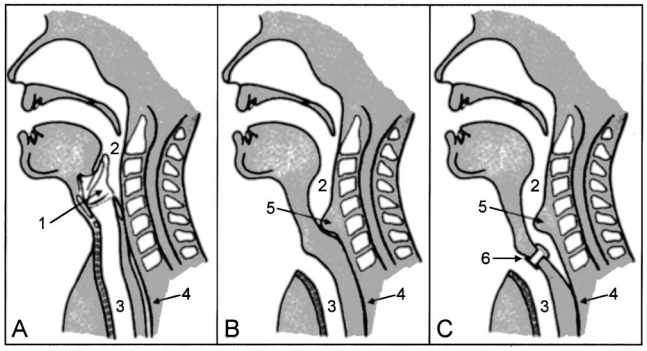


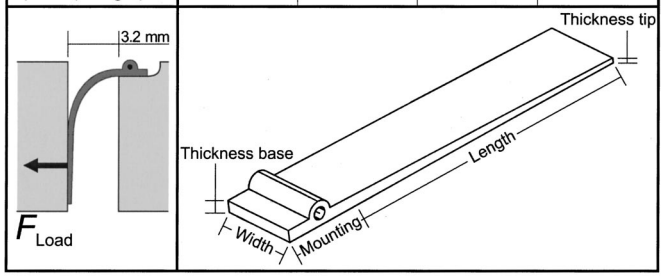
FIG. 1. (A) Head and neck region anatomy before laryngectomy; (B) complete separation of airway and digestive tract after laryngectomy; (C) laryngectomee with shunt valve situated in tracheoesophageal fistula. 1, Larynx; 2, pharynx; 3, trachea; 4, esophagus; 5, pharyngoesophageal segment; 6, tracheoesophageal shunt valve.

ciple (e.g., flutes) is the dependency of f_0 on vocal tract configuration. The principle of the siren was rejected because its rotating part appeared to be too sensitive to tracheal secretions. Inward-striking reeds (e.g., woodwind-type reeds) were tried before, but could not vary f_0 sufficiently.[12,13] An outward-striking lip reed[14] was selected, which resembles the cyclic opening and closing of the gap between the lips of a musician playing a brass instrument.[15] The new voice source consists of a single floppy lip, which performs self-sustaining flutter-type oscillations driven by the expired pulmonary air that flows through the TE shunt valve along the outward-striking lip. In the neutral position, the straight silicone rubber lip reed is bent 90° against the opposite wall to obtain the load force necessary for proper functioning (Table I).

A prototype of this artificial voice source only proved beneficial in female laryngectomees with a hypotonic PE segment.[16] For them the artificial voice source improves voice quality and increases the average pitch of voice, without decreasing intelligibility or necessitating other pressure and airflow rates than regular TE shunt valve speech. Patients can regulate pitch and sound intensity of this prosthetic voice to a limited extent by altering the airflow along the floppy lip. The mechanism does not result in unacceptable airflow resistance. However, for this new mechanism of postlaryngectomy voice restoration to become useful for a larger group of laryngectomees, a voice with an f_0 suitable for males has to be generated as well. In addition, a larger part of the melodic and dynamic range of the sound source has to be reached within the regular airflow range for TE shunt valve speech.

TABLE I. Dimensions and physical properties of the evaluated lip configurations (length=12.5 mm; width=3.3 mm).

Lip configurations	Thickness base (mm)	Thickness free tip (mm)	Load force (10^{-4} N)	Natural freq. unbent (Hz)
Lip 7-4 (tapered)	0.7	0.4	67	29
Lip 7-3 (tapered)	0.7	0.3	66	27
Lip 6-4 (tapered)	0.6	0.4	32	24
Lip 6-3 (tapered)	0.6	0.3	44	25.5
Lip 5-3 (tapered)	0.5	0.3	18	19
Lip 5-2 (tapered)	0.5	0.2	15	18
Lip 4-2 (tapered)	0.4	0.2	7	16
Lip 3-3 (straight)	0.3	0.3	4	11
Lip 3-2 (tapered)	0.3	0.2	4	11
Lip 2-2 (straight)	0.2	0.2	1	6.5



To determine the optimum configuration of the voice source, De Vries *et al.* developed a numerical model.[17] In this model, the Finite Element Method is used to describe the geometry and mechanical properties of the floppy lip reed in the prosthesis. The airflow through the voice source is described by two-dimensional Navier–Stokes equations. The interaction between the floppy lip and the airflow is modeled by placing both in a collective grid. This model however, is just two dimensional, thus not allowing for possible torsion of the floppy lip and leakage of air along the lip. Moreover, stroboscopic observations show a closed phase in the lip cycle, whereas in the simulations no contact was observed between the lip and the opposite wall. These limitations cause significant differences between the simulated and measured lip motion characteristics, thereby underscoring the need for actual observational data.

The purpose of this study is to obtain aero-acoustic measurement data and detailed high-speed photographic sequences of several lip configurations, on which future floppy lip reed models may be based. Visualization may provide a more explicit understanding or reveal features of lip reed behavior not previously known.[18] The aero-acoustic measurement data allow us to appreciate the effect of lip configuration on the effort in terms of expiratory airflow and tracheal pressure, required from patients to attain specific vocal performance (i.e., intensity and f_0) and voice quality parameters (i.e., harmonics-to-noise ratios). Additionally, these experimental data can be used as quantitative classification criteria to establish the most promising lip configurations for clinical use.

II. EXPERIMENTAL METHODS

A. The silicone rubber lip reeds

Implant-grade liquid silicone rubber was chosen for several reasons: (a) the biocompatibility of the lip has to comply with the FDA Code of Federal Regulations, title 21, part 874.3730; (b) a silicone rubber lip reed has a lower modulus of elasticity than, for example, metal reeds, thus enabling lower fundamental frequencies and requiring less airflow; (c) the movement of a vibrating flexible lip can be arrested very abruptly during each cycle by striking against its housing, yielding sound with a broad spectrum of harmonics, like laryngeal voices, which is likely to improve naturalness of the artificial voice;[19] and (d) liquid silicone rubber can be injection molded under under vacuum and cured at room tempera-

ture in almost any shape of mold. The molds were engineered using the CAD/CAM package I-DEAS® (SDRC, Milford, OH, USA).

A previous study indicated that twin-lipped configurations generate an unnaturally high f_0 compared to single-lipped configurations,[11] which probably results from a higher load force and from the more confined space for the oscillatory cycles of their twin lips. Moreover, analogous to human vocal folds, two opposing lips have to be mounted exactly symmetrical in order to function properly, which proved very difficult to reproduce.

Prior to the current study, we explored the aero-acoustic characteristics of a wide range of lip configurations *in vitro*, aiming at a voice with an f_0 suitable for male laryngectomees. This showed us that bedding metal or silicone rubber ballast in the tip of these floppy lip reeds is not an appropriate method to decrease f_0, since it made them prone to chaotic behavior and required high airflows for the onset of sound production. Also decreasing the load force on the lip reed, either by providing them with crosswise grooves, or by increasing the distance to the opposite canal wall, proved unsuccessful. It reduced the intensity of the generated sound and again made the lip prone to chaotic behavior. Just making larger lips for lower frequencies is not possible, because the dimensions of the lip are confined by the geometry of the silicone rubber shunt valve that is determined by the size of the TE fistula. To facilitate lower fundamental frequencies for male laryngectomees, the lips evaluated in the current study consist of softer silicone rubber and are longer than previous prototypes,[11,16] which ranged from 6.5 to 10.5 mm. The durometer of the employed silicone rubber is 43 Shore A units, the density is 1130 kg/m^3 and the elasticity modulus at 200% elongation is 3.45×10^6 Pa.

The results obtained from the numerical model indicated that tapered lip configurations might have a relatively low threshold of vibration and a relatively steep increase of f_0 with increasing pressure and flow.[17] These two features are important for patients since they determine, respectively, the effort required for the onset of phonation and the potential for natural prosody using the prosthesis. Therefore, the lip configurations evaluated in this study are mostly tapered. To investigate the effects of lip reed stiffness, load force, and mass distribution on aerodynamic requirements and acoustic performance of the sound source, a set of 10 silicone rubber lips having the same length and width ($12.5\,\text{mm}\times3.3\,\text{mm}$), but varying in thickness and shape, was employed (Table I). The lip configurations are named after the thickness of their base and free tip, i.e., tapered lip 5-3 converges from 0.5 mm at the base to a free tip of 0.3 mm. The natural frequency of each unbent lip reed and the load force exerted by bending it 90° against the opposite wall are also given in Table I.

A perspex setup to determine the aero-acoustic characteristics of the lip reeds *in vitro* and obtaining the high-speed sequences is depicted in Fig. 2. The ringed tracheal tube is connected to a high-pressure air supply through a pressure regulator. The dimensions of the canal between the tracheal and esophageal side of the model are identical to the dimensions of the square lumen in the shaft of the TE shunt valve that is developed for use in patients: a height of 3.2 mm and a width of 3.5 mm, allowing for 0.1 mm clearance at both sides of the lip. The lip is mounted in this canal by means of a stainless steel rod that is encased in the base of the lip at the time of molding.

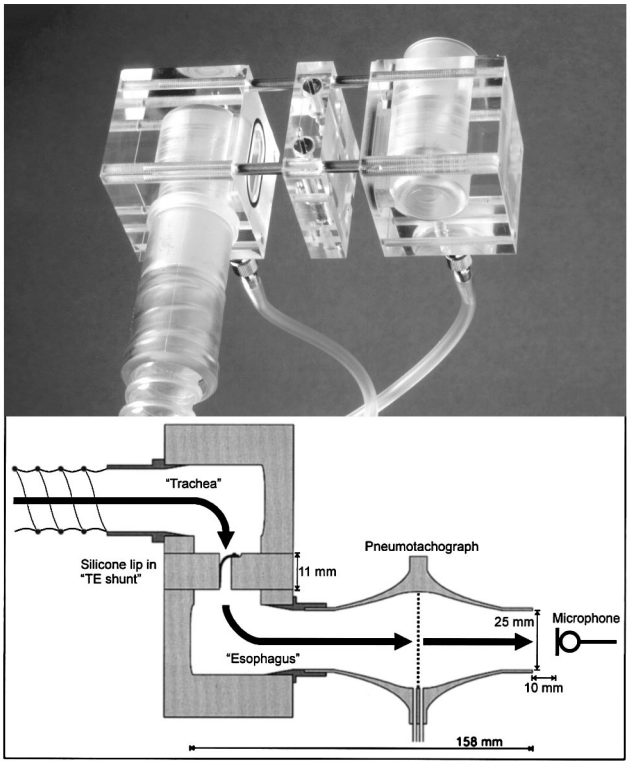


FIG. 2. Experimental setup representing a tracheoesophageal fistula. *Top*: a photograph of the opened perspex setup, showing the 'tracheal' side of the model on the left, the 'esophageal' side on the right, and the canal containing the lip reed in between. *Below*: a schematic cross-section of the experimental setup.

B. Aerodynamic and acoustic data acquisition

A differential pressure transducer (DT-2-14P-0-10L, Modus Instruments, Clinton, MA, USA) measured the static air pressure drop (P_0-P) across the lip. The volume flow through the device was determined downstream using a respiratory flowhead (F300L, Mercury Electronics, Glasgow, Scotland), connected to another differential pressure transducer (DT-2-02P-0-10L, Modus Instruments). These transducers were calibrated by means of a calibration analyzer (Timeter RT-200, Allied Healthcare, St. Louis, MO, USA). A miniature electret condenser microphone (9468, Microtronic, Amsterdam, The Netherlands) with a hybrid integrated amplifier was positioned 10 mm downstream of the sound outlet of the flowhead. The microphone was connected to an additional amplifier (SCXI-1120, National Instruments, Austin, TX, USA). A digital A-weighted sound level meter (CEL-231, CEL Instruments, Hitchin, England) was positioned 0.30 m downstream from the wire screen of the flowhead, analogous to our standard mouth-microphone distance for measurements *in vivo*. The four sensor signals were sampled simultaneously at 10 000 samples/second by a PC based 16-bit AD converter (National Instruments PCI-MIO-16xe-10) and digitally processed to byte stream files by custom-built data acquisition software (National Instruments LabVIEW®

5.1). Low-pass eighth-order elliptic filters (National Instruments SCXI-1141) were applied to all sensor signals at 3500 Hz to prevent aliasing.

Each of the of lips was recorded twice while the volume flow increased slowly from zero to 4×10^{-4} m^3/s and *vice versa*, taking approximately 60 seconds per recorded loop. For every 100 ms the fundamental frequency was calculated off-line from the recorded microphone signals by another custom-built LabVIEW® application, based on a harmonic product spectrum algorithm.[20] This robust algorithm not only uses the fundamental but also the positions of the higher harmonics in the spectrum by calculating the largest common divider of their frequencies.

The simultaneously registered signal of the sound level meter was used to select specific periods of the recorded loops: each time the meter crossed 60 dB[A] (onset), 70 dB[A] (regular conversation), 80 dB[A] (loud conversation) or 90 dB[A] (shouting), a sample of 1000 ms was cut from the recording to be used for acoustic analysis. Since each lip was recorded twice, both *crescendo* and *decrescendo*, this resulted in four samples at each sound pressure level, adding up to 16 samples per lip. The level of noise in these 1000 ms samples was estimated with a cepstrum-based technique, described by De Krom for laryngeal voices.[21] Harmonics and noise of a voice spectrum are disentangled in the corresponding cepstrum (the Fourier transform of the log spectrum) and can then be manipulated separately. The harmonic structure of the spectrum gives rise to a few equidistant peaks in the cepstrum. The noise affects various cepstral aspects: the spectral envelope of the noise contributes to the low end of the cepstrum, whereas its fine structure defines the higher regions of the cepstrum. By application of a comb filter to the cepstrum, energy related to the harmonic structure in the spectrum can be removed. After a inverse transformation and a level correction, an estimate of the noise spectrum is obtained. By subtraction of the estimated noise spectrum from the original voice spectrum, the harmonics-to-noise ratio was calculated for two frequency regions (<700 Hz and 700–2300 Hz). These regions roughly correspond to the frequency regions of first and second formant.[22] To obtain one set of signal parameters per lip at each sound pressure level, outcomes of calculations on four samples corresponding to identical sound pressure levels were averaged.

C. High-speed digital imaging

A magnifying laryngo-pharyngoscope by Stuckrad (4450.57, Richard Wolf GmbH, Knittlingen, Germany) equipped with a fiber light cable that provided coaxial light from a 300 Watt xenon lamp (Auto-LP/Flash 5135, Richard Wolf GmbH) was positioned right above the canal between the tracheal and esophageal side of the clear perspex setup. On the endoscope we mounted a Weinberger SpeedCam Pro monochrome camera (Weinberger AG, Dietikon, Switzerland). The CCD image sensor of this high-speed digital imaging system has a square pixel array of 512×512 pixels, which can operate up to 1000 frames per second at full resolution. By reducing the horizontal resolution to 512×128 pixels we were able to increase the recording speed to 3000 frames per second of 8-bit gray scale. Expansion of the internal memory of the camera to 1024 megabytes provided for 4 seconds recording time (storage of 12 000 images at the applied resolution). Each of the lips was recorded for 4 seconds, while the volume flow increased from zero to 4×10^{-4} m^3/s. To avoid storing gigabytes of images, several distinctive sequences of 200 pictures (66.7 ms) were selected per lip for storage as manageable Microsoft Windows AVI files.

III. EXPERIMENTAL RESULTS

A. Aerodynamic and acoustic analyses

Figure 3 shows the volume flow, the sound pressure level measured 0.30 m downstream from the wire screen of the flowhead, and the fundamental frequency of lip 5-3 (Table I) plotted versus the pressure difference (P_0-P) across the lip. Similar data were obtained for all lips.

Pressure difference, volume flow, and f_0 of all lips are presented in Table II at four selected sound pressure levels. The harmonic product spectrum algorithm was not able to determine the f_0 in the highly irregular signals representing lip 3-3 at 70 dB[A] and lip 6-4 at 90 dB[A]. No lip configu-

TABLE II. Pressure difference (P_0-P) across the lip, volume flow through the device, and fundamental frequency per lip for each sound pressure level. Desired f_0 values for male laryngectomees are italic.

	60 dB[A]			70 dB[A]			80 dB[A]			90 dB[A]		
Configuration	P_0-P (10^3 Pa)	Flow (10^{-4} m^3/s)	f_0 (Hz)	P_0-P (10^3 Pa)	Flow (10^{-4} m^3/s)	f_0 (Hz)	P_0-P (10^3 Pa)	Flow (10^{-4} m^3/s)	f_0 (Hz)	P_0-P (10^3 Pa)	Flow (10^{-4} m^3/s)	f_0 (Hz)
Lip 7-4	0.60	0.40	152	1.20	1.20	170	1.90	1.70	200	5.30	3.20	260
Lip 7-3	0.55	0.30	154	1.25	1.40	172	1.70	1.70	202	4.80	3.30	264
Lip 6-4	0.45	0.30	138	0.70	1.00	*145*	1.70	1.90	197	4.65	3.40	[a]
Lip 6-3	0.65	0.80	*118*	0.85	1.10	*150*	1.65	1.60	*180*	4.30	3.40	*245*
Lip 5-3	0.45	0.70	*115*	0.55	0.80	*143*	1.35	1.60	*177*	4.30	3.70	*240*
Lip 5-2	0.45	0.80	*102*	0.75	1.10	*150*	1.85	2.10	*168*	3.75	3.50	*238*
Lip 4-2	0.40	1.00	*105*	0.65	1.20	*146*	1.40	2.20	*160*	3.50	4.00	*210*
Lip 3-3	0.25	0.70	*90*	0.55	1.30	[a]	1.05	1.90	85	3.35	3.70	*180*
Lip 3-2	0.30	0.90	*100*	0.55	1.40	90	1.05	2.10	*140*	2.55	3.40	400
Lip 2-2	0.20	0.80	75	0.48	1.50	95	1.00	2.20	125	90 dB[A] not attained		

[a]The harmonic product spectrum algorithm was not able to determine f_0 in the highly irregular signals representing lip 3-3 at 70 dB[A] and lip 6-4 at 90 dB[A].

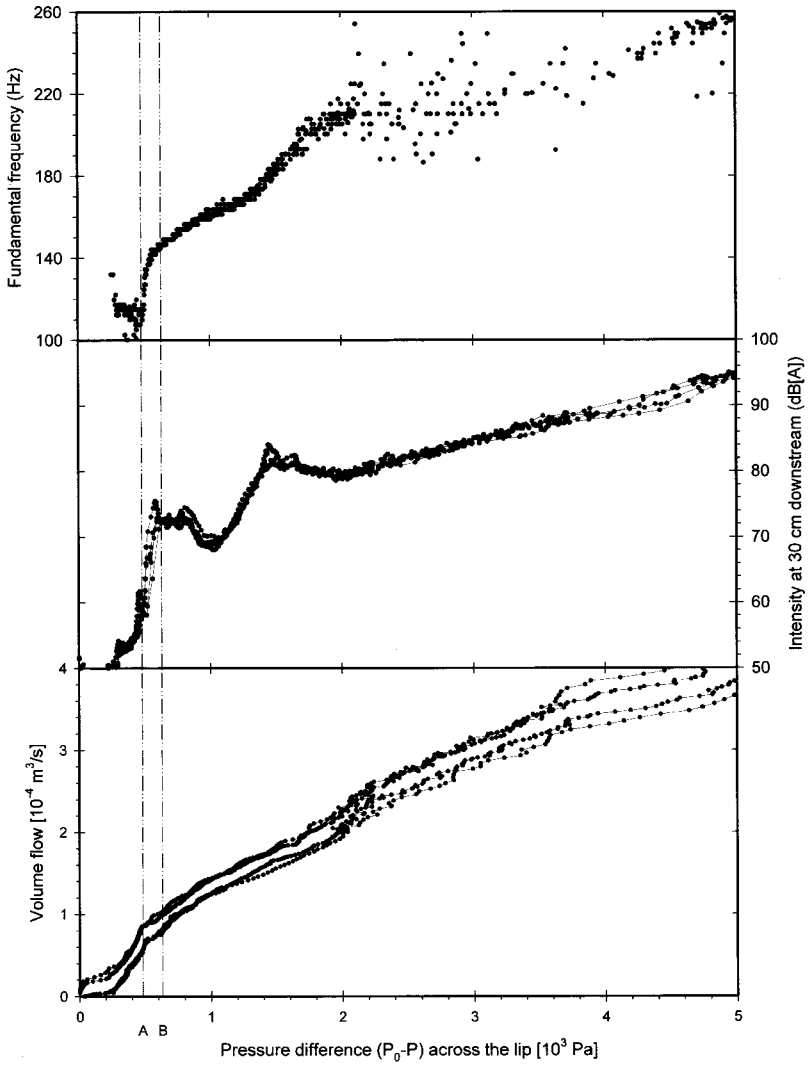


FIG. 3. Volume flow, fundamental frequency, and sound pressure level plotted versus static air pressure drop for lip 5–3. Each point is based on 1000 consecutive samples. Lines A and B refer to the power spectra in Fig. 9.

ration requires more than 1×10^{-4} m^3/s airflow for onset of sound production. At regular conversational intensity, 70 dB[A], the lips require approximately 1.2×10^{-4} m^3/s airflow. For most lips this increase of volume flow yields a higher f_0 than at 60 dB[A]; in lips 3-3, 3-2, and 2-2 however, f_0 initially decreases 10–20 Hz. At 80 dB[A], loud conversational intensity, the lips require approximately 1.9×10^{-4} m^3/s airflow. For all lips this increase of volume flow yielded a higher f_0 than at 70 dB[A]. All lips except the thinnest straight lip (lip 2-2) are able to generate 90 dB[A], requiring approximately 3.6×10^{-4} m^3/s airflow.

At each sound pressure level, f_0 and corresponding pressure difference increase with lip thickness, ranging from 75–155 Hz and $2-6\times10^2$ Pa, respectively, at onset, to 180–265 Hz and $3.4-5.3\times10^3$ Pa at 90 dB[A]. Lip 3-2 however, shows a sudden increase of f_0 [from 185 Hz to 370 Hz (these frequency jumps being octaves seems purely accidental; we have neither a theoretical explanation for this, nor any clue from the high-speed digital image sequences)] upon exceeding a volume flow of 2.8×10^{-4} m^3/s. The same irregular behavior (f_0 increase from 280 Hz to 560 Hz) is demonstrated by lip 6-4 at a volume flow of 3.9×10^{-4} m^3/s. For both lips this phenomenon is accompanied by a 20% increase of the airflow resistance of the device. Most of the other lip configurations show evidence of this behavior as well, but at a volume flow higher than 4×10^{-4} m^3/s.

Several characteristics of voice production in normal male laryngeal speakers and laryngectomees using a regular TE shunt valve, are presented in Table III. This table is restricted to only a few parameters, important for our purpose, and summarizes what appear to be accepted reference values. Within the regular airflow range for TE shunt valve speech, $0.5-3.0\times10^{-4}$ m^3/s, all lips are able to generate at least 85 dB[A] (Table IV). Thin-based lips however, operate at a smaller static pressure difference at this intensity than thick-based lips that require approximately $3-5\times10^3$ Pa. The potential for patients to obtain natural prosody by means of this prosthetic voice can be estimated by considering the range of fundamental frequencies generated by each lip configuration within the physiological airflow range (Table IV). Most lips offer a range in the order of 100 Hz. Lips 4-2, 5-2, 5-3, and 6-3 approximate the mean f_0 range for male laryngeal speakers best. In Tables II and IV, f_0 values matching the reference values of male laryngeal voices are shaded gray.

The cepstrum-based technique, used to estimate the harmonics-to-noise ratios in these samples, depends on correct f_0 extraction. Since our algorithm was not able to determine the f_0 in the highly irregular signals representing lip

TABLE III. Reported characteristics of voice production in normal male laryngeal speakers and laryngectomees using a regular TE shunt valve.

Physiological voice characteristics	Laryngeal voice	TE fistula voice
Phonatory volume flow	mean[a]: 1.9	median: 1.0
(10^{-4} m^3/s) (Refs. 34, 35)	range: 0.9–3.4	i.q. range[b]: 0.5–3.0
Static tracheal pressure	mean[a]: 0.6	mean[c]: 6.9
(10^3 Pa) (Refs. 7, 35)	range: 0.3–1.5	range: 4.2–10.9
Max. sound pressure level	mean: 117	mean: 71
(dB) (Refs. 5, 36)	range: 105–126	range: 66–77
Speaking fundamental frequency	mean[d]: 124	mean: 104
(Hz) (Refs. 8, 23)	σ[a]: 21	σ: 18
Melodic range[f] (Hz)	mean: 595	mean: 84
(Refs. 5, 37)	σ: —	σ: 32
Harmonics-to-noise ratio	median: 25	median: 4
(dB) below 700 Hz (Ref. 23)	range: 22–31	range: 1–9

[a]Mean for 25 male speakers, averaged over 15 syllables at normal intensity.
[b]i.q. range=interquartile range.
[c]Intratracheal pressure for sustained phonation at comfortable intensity, averaged over 54 laryngectomees.
[d]Average speaking f_0 of 1003 male adults found in the literature.
[e]σ=standard deviation.
[f]Total fundamental frequency range, or phonational range.

3-3 at 70 dB[A] and lip 6-4 at 90 dB[A], no harmonics-to-noise ratios could be calculated for these two conditions either.

At 90 dB[A], the harmonics-to-noise ratios of all lips deteriorate (Table V). At conversational intensities (70 and 80 dB[A]), the harmonics-to-noise ratios in the frequency regions containing the first and second formant are appreciably good (>25 dB and >10 dB, respectively) in the three thickest tapered lips 7-4, 7-3, and 6-4. On the other hand, the harmonics-to-noise ratios of the four thinnest lips 4-2, 3-3, 3-2, and 2-2 are relatively low. Using the same algorithm, Festen *et al.*[23] found the harmonics-to-noise ratio in the frequency region of the first formant to range from 22 to 31 dB in normal laryngeal voices, and from 1 to 9 dB in laryngectomized TE shunt valve speakers. Of the lips selected above for their optimal f_0 range, lip 5-3 generates considerably stronger harmonics than the others, both in the region of the first and second formant.

B. High-speed digital imaging

The sequences showing the onset of vibration are much alike for all lips, although the initial amplitude appears to be larger in the thinnest lips. Figure 4 shows several phases of the onset of lip 5-2, starting at $t=0$ ms with the lip in its resting position abutting the opposite wall just before initial movement [Fig. 4(A)]. After 8.7 ms, 26 images later, the increasing overpressure at the tracheal side of the lip has gradually pushed the lip to its first peak opening [Fig. 4(B)]. The recoil of the lip reed and the Bernoulli force, a local minimum of pressure created by the flow through the device, causes the narrowest (middle) part of the valve channel to collapse rapidly, leading to the first closure at $t=10.7$ ms. After the first closed phase, lasting 2.7 ms, the lip opens again to let the second flow pulse pass between $t=13.3$ ms and $t=21.3$ ms. The lip centerline was traced by hand separately for each digital image taken during this period and then compiled to show the entire sequence [Fig. 4(C)]. This motion is essentially one of wave propagation from the base to the tip of the lip, with a one-sided constraint at the opposite wall, for which a lip-valve model has been discussed by Ayers.[24] For thin lips the transition from the initial cycle to a full-amplitude self-sustaining oscillation requires only a few cycles, whereas thick lips build up amplitude more gradually over a considerable number of cycles.

Contrary to the predictions based on the numerical simulation model,[17] contact was observed between the lips and the walls of the housing. Immediately after onset, this contact takes place twice per cycle at closure. One after the other, the mid-part of the lip and the free tip bump gently into the opposite wall [Fig. 4(C)]. At this stage the level of sound produced is very low. When increasing the volume flow, the free tip additionally strikes the wall on the side where the lip is mounted at the time of maximum opening.

TABLE IV. Maximum intensity, highest fundamental frequency, and lowest fundamental frequency per lip within regular airflow rates for TE shunt valve speech.[34] Desired f_0 values for male laryngectomees are italic.

	Maximum intensity		Highest f_0		Lowest f_0	
Configuration	P_0-P (10^3Pa)	SPL (dB[A])	P_0-P (10^3Pa)	f_0 (Hz)	P_0-P (10^3Pa)	f_0 (Hz)
Lip 7-4	4.75	88	4.75	252	0.70	144
Lip 7-3	4.10	89	4.45	260	0.65	146
Lip 6-4	4.00	87	3.40	*236*	0.50	130
Lip 6-3	3.70	86	3.70	*220*	0.50	*118*
Lip 5-3	3.10	85	3.10	*222*	0.40	*115*
Lip 5-2	2.55	87	3.05	*213*	0.35	*90*
Lip 4-2	2.35	87	2.35	*180*	0.20	*85*
Lip 3-3	2.35	87	2.35	160	0.15	70
Lip 3-2	2.25	86	2.25	390	0.15	65
Lip 2-2	1.30	87	1.70	170	0.15	73

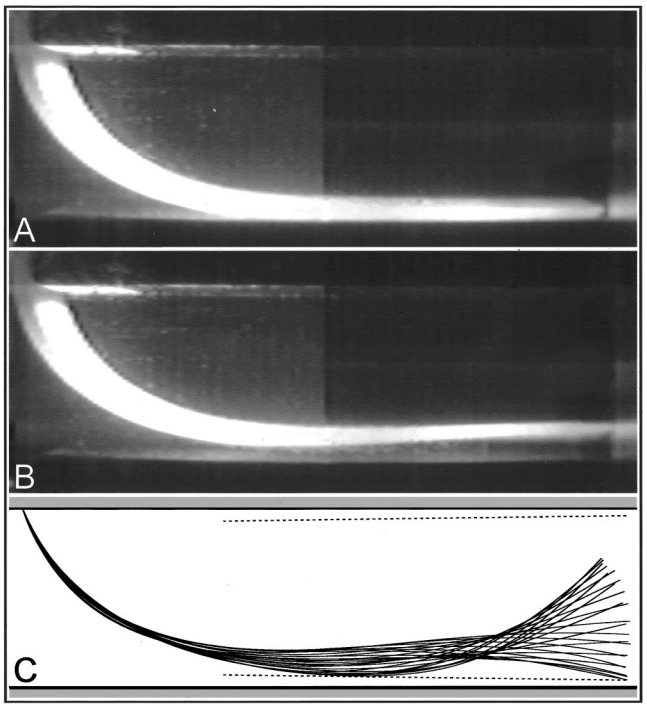


FIG. 4. Onset of vibration for lip 5–2, showing (A) the lip in neutral resting position at t=0 ms, (B) the first point of maximum opening at $t=8.7$ ms, and (C) a tracing of the second cycle of the lip between $t=13.3$ ms and $t=21.3$ ms. Where the traced lip centerline coincides with the dotted line, the tapered lip comes into contact with the housing.

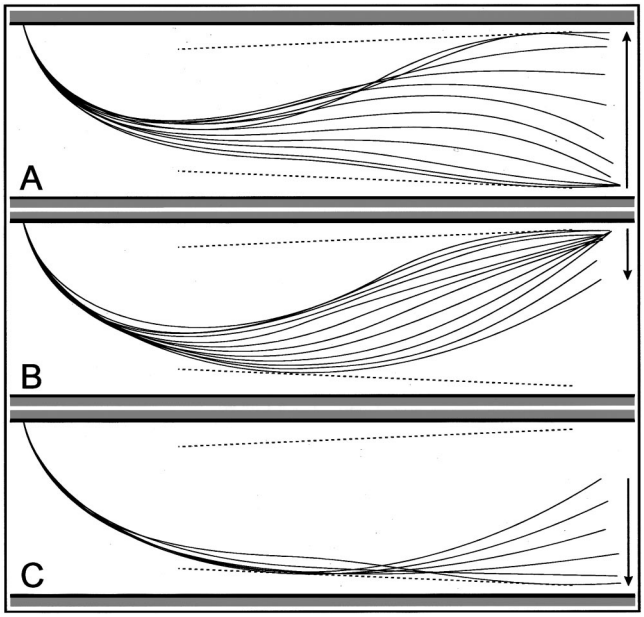


FIG. 5. A tracing of an entire cycle of lip 5–3 during steady-state vibration at 112 Hz. Where the traced lip centerline coincides with the dotted line, the tapered lip comes into contact with the housing. Images 1–12 (A) show the major opening phase of the cycle, images 12–23 (B) show the major closing phase of the cycle, and images 23–27 and 1 (C) show the phase in between the two closures of this cycle.

This probably perturbs the lip motion in such a way that the airflow variations in the artificial voice source are increased, giving rise to a steep increase of the generated sound intensity. The transition from two-contacts behavior to three-contacts behavior shows hysteresis.

Figure 5 shows a tracing of a 27-image sequence of an entire cycle of lip 5-3 during steady-state vibration at 112 Hz. Again, the lip centerline was traced by hand separately for each digital image and then compiled to show the entire sequence. By dividing the entire cycle in three phases, the three-contacts behavior described above is illustrated distinctly. A similarly complex, though regular, flutter-type motion pattern can be seen in all lips for some part of their melodic and dynamic range.

From the high-speed digital image sequences, it is apparent that the free tip striking against the housing induces a small superimposed traveling wave in the lip that starts at the tip and runs toward the base of the lip. This disturbance appears to happen especially when the lip strikes against the wall on the side where the lip is mounted, at the time of maximum opening. This effect is much more noticeable when playing the sequences slowly than when studying the individual frames. In thick stiff lips, the amplitude of the traveling wave is small and its propagation fast, which gives the impression of a shudder of the lip upon striking against the wall. In thin floppy lips however, the amplitude is higher and the propagation slower, which enables us to see the small superimposed wave run from the tip to the base of the lip and often back and forth again.

As expected, the increased stiffness and higher load force of thick-based lips yield higher frequencies compared to thin lips. The mean opening area is significantly smaller for these lips, corresponding with the finding of higher airflow resistance in these lip configurations. Moreover, thick lips show the same regular flutter-type motion pattern at almost any volume flow; their stiffness hardly allows for cycle-to-cycle irregularities. Thin lips on the other hand, may demonstrate chaotic behavior such as torsion of the body of the lip [Fig. 6(A)], torsion combined with superimposed traveling waves [Fig. 6(B)], or curling of the free tip [Fig. 6(C)], even at a moderate volume flow.

IV. DISCUSSION

In this experimental study, aero-acoustic data and detailed high-speed photographic sequences were obtained for 10 lip reed configurations. The high-speed visualizations of lip reed behavior are useful because of both the overall insight and the details they provide. Details of lip reed behavior may be crucial for both understanding[18] and optimizing the aero-acoustic characteristics of this artificial voice source. Next, we discuss several aspects of lip reed behavior in general and implications for the use of the artificial voice source.

A. Experimental setup

The acoustic properties of the model upstream and downstream from the lip do not reflect the conditions *in vivo*. However, Van der Plaats *et al.*[25] found that the impedance spectrum of the subglottal tract, i.e., trachea, bronchi, and

TABLE V. Harmonics-to-noise ratio (dB) in the frequency regions of the first formant (*F*1) and the second formant (*F*2) per lip for four sound pressure levels. Favorable harmonics-to-noise ratios for laryngectomees are italic.

	60 dB[A]		70 dB[A]		80 dB[A]		90 dB[A]	
Configuration	*F*1	*F*2	*F*1	*F*2	*F*1	*F*2	*F*1	*F*2
Lip 7-4	*15.9*	*8.8*	*23.7*	*7.7*	*25.1*	*12.7*	*12.6*	*5.5*
Lip 7-3	*29.5*	*20.5*	*29.9*	*14.6*	*25.0*	*13.9*	6.8	2.4
Lip 6-4	*28.4*	*25.8*	*27.7*	*16.3*	*27.2*	*11.5*	no f_0 determined	
Lip 6-3	6.0	0.4	*14.2*	4.3	*11.2*	2.0	6.8	*5.6*
Lip 5-3	*21.2*	4.7	*28.6*	*15.3*	*18.7*	*7.2*	6.6	3.0
Lip 5-2	*14.6*	3.8	*21.6*	*9.5*	*10.5*	2.2	5.6	2.7
Lip 4-2	*13.9*	4.4	*12.5*	4.6	1.7	0.6	1.9	0.8
Lip 3-3	6.3	1.1	no f_0 determined		1.5	0.3	1.1	0.5
Lip 3-2	8.2	1.7	0.8	0.3	2.2	0.6	6.2	*6.7*
Lip 2-2	3.6	0.7	2.2	0.2	1.5	0.2	90 dB[A] not attained	

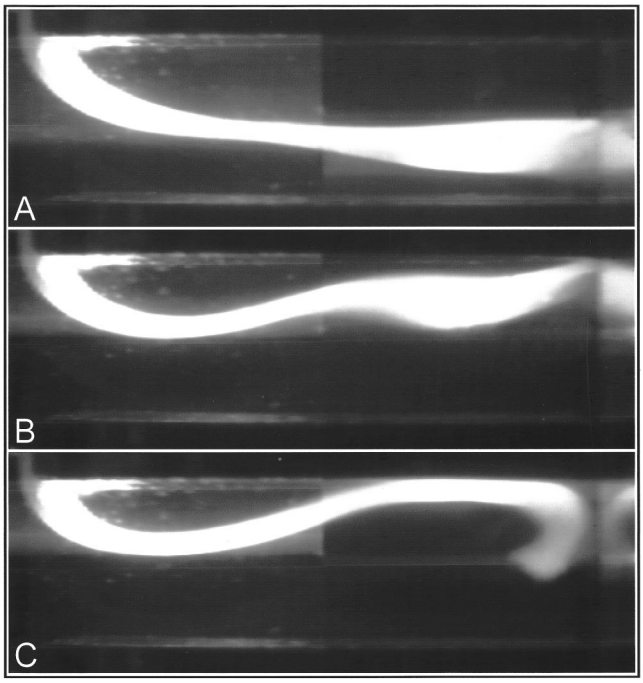


FIG. 6. Chaotic behavior of lip 4–2 at approximately 75 dB[A], such as torsion of the body (A), torsion combined with superimposed traveling waves (B), and curling of the free tip (C). Rapid motions of the free tip cause partial blurring of the lip contours in these stills.

lungs, does not affect the output of this alternative voice source based on an outward-striking lip reed. Nevertheless, we repeated some of the experiments described above after (a) augmenting the length of the downstream vocal tract to an anatomically correct 0.17 m,[26] and (b) installing sound-absorbent porous foam 0.20 m upstream in the ringed tracheal tube to simulate the acoustic impedance of the lungs.[27] These simple modifications, reflecting the acoustic load working on the glottis *in vivo*, had no effect on any of the relations between volume flow, static air pressure drop, f_0, and sound pressure level, thus supporting the observation by Van der Plaats *et al.*[25] and the validity of our measurements.

Differently from conditions *in vivo*, where the artificial voice source is incorporated in a silicone rubber TE shunt valve, the lips are mounted in a valveless perspex model. Although the mounting method and the dimensions of the canal between tracheal and esophageal side of the experimental setup are identical to *in vivo* conditions, the setup does not include a shunt valve, since we are primarily concerned with visualizing and understanding the characteristics and behavior of the lip reed proper. Incorporating the artificial voice source in future TE shunt valves will yield a higher airflow resistance compared to this valveless situation[11,16] and possibly a modulation of the generated tones. On the other hand, the model contains several sharp edges, such as the abrupt inlet of the tracheoesophageal canal (Table I and Fig. 2), which may induce a *vena contracta* effect.[28] This artifact of the experimental setup would probably result in higher airflow resistance compared to the smooth human anatomy.

B. Determinants of f_0: volume flow, lip length, and lip base thickness

Laryngectomees using this new voice source require a control parameter to vary f_0 in order to obtain a functional intonation pattern. Likewise, engineers of future sound-producing shunt valves have need of design parameters to vary the mean speaking f_0 in order to match the gender-specific reference values of laryngeal voices. The current study yields both.

In the human vocal system, intensity and frequency are not controlled independently. Laryngeal speakers tend to raise their voice in pitch when they raise their voice in loudness, and they do it differently in different portions of their vocal range.[29] For the chest register, Titze[30] demonstrates that the typical rise in f_0 with subglottal pressure observed in human and canine larynges (2–6 Hz per 10^2 Pa) is explainable on the basis of an amplitude-dependent tension in the vocal folds. This not only explains intensity and frequency being coupled, but also subglottal pressure being a variable of control for both.

Also in the evaluated silicone rubber lip reeds, an almost linear relation is found between f_0 and the static pressure difference across the lip for the largest part of their effective melodic and dynamic range, as shown in Fig. 3 for lip 5-3. Aurégan and Depollier[31] report a similar behavior for both a numerical and an *in vitro* model of the soft palate during snoring. Despite complex flow characteristics (dynamic flow separation, vortex formation, etc.), f_0 of the soft palate motion increases almost linearly with air velocity, when multiple contacts per oscillation between the soft palate and the pharyngeal walls occur. Actually, Hirschberg *et al.*[32] state that elastic beating of the reed on the lay of reed organ pipes explains the dependence of f_0 on blowing pressure in these woodwind instruments. Hydrodynamic effects are significant when the reed does not beat. If beating occurs, however, the hypothetical trajectory of the reed is abruptly clipped by its collision with the lay. In Hirschberg's mass-spring model for this phenomenon, the resonance frequency increases with increasing blowing pressure. Therefore contact between the lip and the walls of the housing seems to be a prerequisite for attaining a wide f_0 range. This melodic range, in the order of 100 Hz within the regular airflow range for TE shunt valve speech, may offer laryngectomees better possibilities for natural prosody than previous prototypes.[11,16] Nonetheless, even these new lips will probably require thorough practice and individual training to make optimal use of this new means of intonation.

For the three thinnest lip configurations, f_0 initially decreases by 10–20 Hz with increasing volume flow just after the onset of vibration. When increasing the volume flow further, however, f_0 gradually rises again similar to thicker lip reeds. We could not find an explanation for this phenomenon by examining the high-speed digital image sequences of these lips, but the U-shaped relation between f_0 and volume flow was also noted by some of the patients when testing a prototype of the artificial voice source.[16] Thicker lip configurations may demonstrate this behavior as well, but only for an instant, when they still generate less than 60 dB[A],

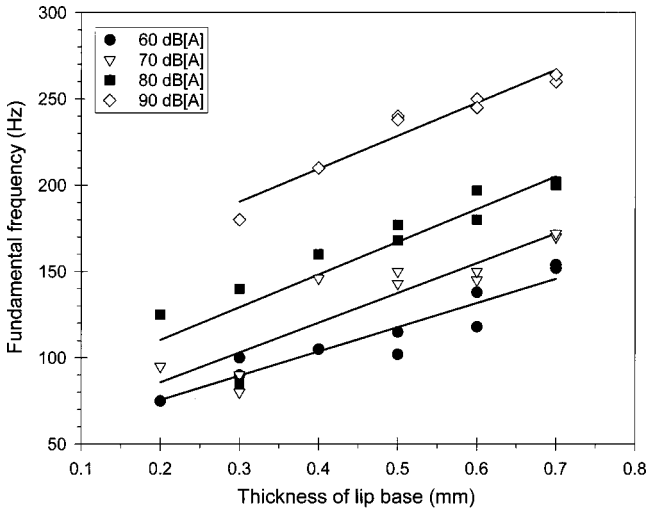


FIG. 7. Fundamental frequency at four sound pressure levels plotted versus the thickness of the lip base in different lip configurations.

which makes the phenomenon hardly audible (Fig. 3, for $P_0-P=0.25-0.45\times10^3$ Pa).

The evaluated lip configurations vary in stiffness, natural frequency, load force, and mass distribution, but not in length. Therefore, no conclusions can be drawn from these data regarding the effect of lip reed length on aerodynamic requirements and acoustic performance of the sound source. However, previous studies indicate that the range of fundamental frequencies generated by an outward-striking lip reed is shifted downward with increasing lip length,[10,11,16] as is also true for freely vibrating cantilevered beams and pipe organ reeds tongues.[33]

In these experiments, the effects of lip reed mass on the frequency of oscillation are confounded with those of lip thickness. This precludes separate conclusions on the effects of lip mass. Since lip mass is directly proportional to thickness, while lip stiffness is proportional to thickness to the third power, it is to be expected that the natural frequency of the unbent lip reed will be proportional to thickness (Table I). Figure 7 presents the frequency dependence on the lip base thickness for the constrained lip reed motion in the current experimental setup. The graph combines results of 160 samples; f_0 was averaged over four measurements at four different sound pressure levels for 10 different lip configurations. Bearing in mind that f_0 is also strongly dependent on the static pressure difference across the lip, the data were fit by a least-squares analysis to four separate straight lines for each of the four different sound pressure levels. These lines fit the data quite well ($r^2=0.87$, 0.82, 0.78, and 0.93, respectively), indicating that the frequency is roughly linearly related to lip base thickness, as is also found for freely vibrating cantilevered beams and pipe organ reeds tongues.[33]

C. Airflow resistance of the artificial voice source

The onset of self-sustaining vibration requires only little driving force in all lip configurations. Thin-based lip reeds, however, initiate flutter-type oscillations at even smaller pressure differences compared to thick-based lip reeds. The extent of the initial horizontal leg in the plot of volume flow

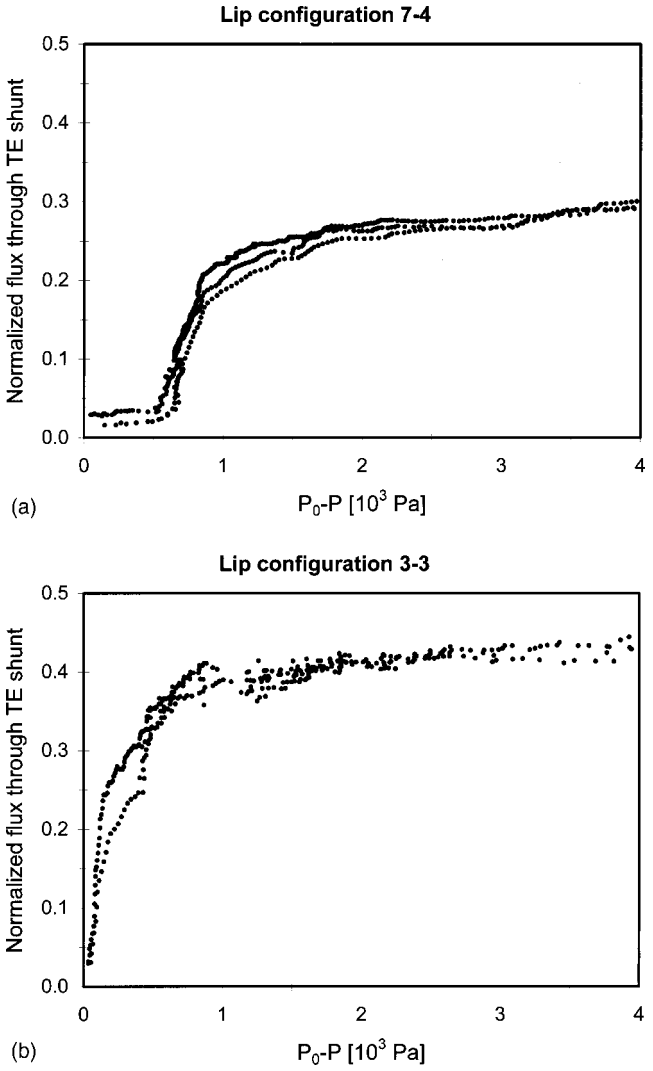


FIG. 8. The dimensionless flux $\Phi/S_{\max}\sqrt{[2(P_0-P)/\rho_0]}$ plotted versus the driving pressure P_0-P for a thick-based lip and a thin-based lip [configurations (a) 7-4 and (b) 3-3, respectively]. The mean air density ρ_0 equals 1.29 kg/m^3 and the area $S_{\max}$ is set to 3.5×3.2 mm.

versus static pressure difference (Fig. 3, for $P_0-P<0.25\times10^3$ Pa) reflects the height of this threshold. Plotting dimensionless flux

$$\frac{\Phi}{S_{\max}\sqrt{\dfrac{2(P_0-P)}{\rho_0}}}$$

versus driving pressure (P_0-P) illustrates this aspect of lip reed behavior more clearly, as shown in Fig. 8. By this definition the normalized flux is the mean volume flow through the valve relative to the volume flow to be expected through the unobstructed lumen. The pressure threshold is almost nonexistent for the displayed thin lip, whereas the thick-based lip requires at least 0.6×10^3 Pa.

Figure 8 also demonstrates the effect of lip base thickness on average cross-sectional area of the reed channel: the normalized flux in the thin lip configuration approaches 0.45, whereas in the thick-based lip configuration the normalized flux approaches only 0.3 at the same static pressure difference. A hypothetical normalized flux of 1 would only be

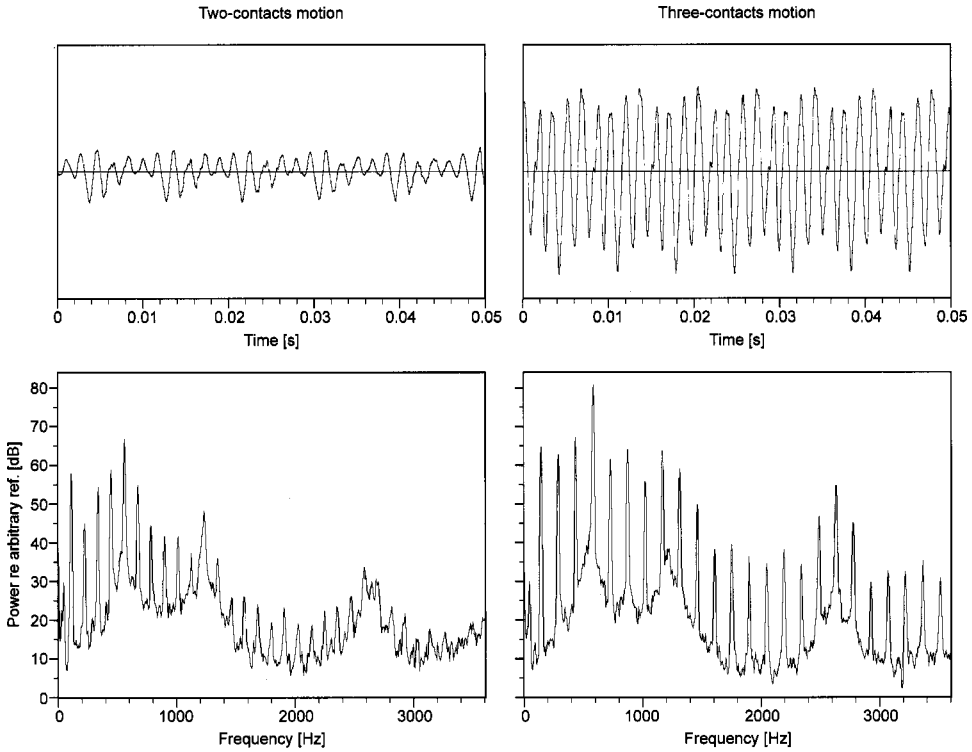


FIG. 9. Temporal and spectral representation of the sound generated by lip 5-3 just before and after its transition from two-contacts behavior to three-contacts behavior (positions A and B in Fig. 3, respectively).

found in a frictionless tracheoesophageal canal without lip reed or inlet discontinuities. The higher load force, brought about by the stiff base of lip 7-4, prevents the lip from bending as far as lip 3-3, resulting in a smaller average cross-sectional area of the reed channel, which is also observed in the high-speed image sequences.

D. Lip reed beating behavior

Contrary to the predictions based on the numerical simulation model,[17] beating of the lip against the walls of its housing occurs over the entire melodic and dynamic range in all lip configurations. Actually, these elastic collisions seem to be essential determinants of lip reed aero-acoustics, not only enabling a wide f_0 range within a limited range of volume flow,[32] but also accounting for most of the observed irregularities in the relations between driving force (P_0-P) on one hand, and both sound pressure level and f_0 on the other. For specific parts of the melodic and dynamic range, different parts of the lip reed (mid-part or free tip) make single or multiple contacts per cycle with different sides of the housing, each combination having its own effect on sound pressure level, f_0, and spectral content. The transitions from one type of behavior to another often show hysteresis.

The effect of the number of contacts per cycle on spectral richness is shown in Fig. 9 for lip 5-3. Just before the transition from the previously described two-contacts to three contacts behavior, at a static pressure difference of 0.48×10^3 Pa (Fig. 3—position A), the harmonics beyond 1500 Hz hardly rise above the level of the noise. Right after the transition, at a static pressure difference of 0.63×10^3 Pa (Fig. 3—position B), when the free tip also beats against the wall on the side where the lip is mounted, the excitation wave form probably is abruptly clipped, yielding stronger harmonics, while the noise level remains the same. The latter behavior is favorable for the artificial voice source, because it results in a less noisy signal in a frequency region that carries the important information on the differences among vowels. Moreover, the spectral slope is less steep after the transition, causing the harmonics beyond 2000 Hz to gain approximately 5 dB relative to the harmonics below 1000 Hz.

The damping of the lip reed, involved in these elastic collisions, is not only determined by the mechanical properties of the silicone rubber, but most likely also by the air cushion trapped between the lip and the wall of the housing. The clearance at both sides of the lip may therefore form an important control parameter of the behavior of beating lip reeds.

E. Harmonics-to-noise ratio

It is interesting to notice that the harmonics-to-noise ratios for thick lip configurations are much better than for thin lips. This may be explained by comparing the acoustic signal in the time-domain and the high-speed digital image sequences of both groups: thin lips are prone to chaotic behavior (torsion, superimposed traveling waves, curling of the free tip, etc.), even at moderate volume flow. At conversational intensities, thick lips not only offer an appreciable harmonics-to-noise ratio in the frequency region that holds the first formant of most vowels, but also in higher regions of the spectrum. The latter may provide good potential for the naturalness of the artificial voice.[19]

At 90 dB[A], the harmonics-to-noise ratio of all lips deteriorates. To roughly discern whether increasing turbulent noise in the experimental setup or a waning harmonic structure is responsible, we measured the sound pressure level of the turbulent noise 0.30 m downstream from the wire screen

of the flowhead at the maximum volume flow of 4×10^{-4} m^3/s, without mounting any lip reed in the perspex model. The level of noise generated by the empty experimental setup is only 51 dB[A], whereas most lip configurations exceed 90 dB[A] at this high volume flow. This supports the hypothesis that, under extreme conditions, even thick lips show chaotic behavior, deteriorating the harmonics-to-noise ratio by breaking up the harmonic structure rather than by increasing the level of turbulent noise.

F. Possibilities of the artificial voice source *in vivo*

The experimental data from this study may be used to find the most promising lip configurations for use in laryngectomees. First, volume flow and static tracheal pressure required for adequate speech levels with this alternative voice source (Table II) have to comply with laryngectomees' physical capabilities (Table III). Second, the mean and range of the fundamental frequency generated within the physiological airflow range (Tables II and IV), need to match the gender-specific reference values of laryngeal voices (Table III). Additionally, voice quality parameters of this prosthetic voice can be considered, such as the harmonics-to-noise ratios (Table V), or the potential of intonation control, reflected in the dependence of f_0 on the static pressure difference across the lip. Since all evaluated lip configurations operate within the limits set by laryngectomees' physical capabilities, the selection of the most promising lip configuration for males will finally be a compromise between the most natural f_0 range for men, generated by thin lips, and the best voice quality obtained from thick-based lips. For females this is easier, since the thick-based lips yield not only a good voice quality, but also generate an f_0 range suitable for women.

V. CONCLUSIONS

We summarize the main conclusions that can be drawn from this study concerning the aero-acoustics of silicone rubber lip reeds for alternative voice production in laryngectomees.

For pressures above the sounding threshold, fundamental frequency, volume flow, and sound pressure level generated by a single floppy lip reed are almost linear functions of the driving force, static pressure difference across the lip. The wide range and the driving force dependence of f_0 can be explained by the elastic beating of the lip against the walls of the housing.

Observed irregularities in these relations are mainly caused by transitions from one type of beating behavior of the lip against the walls of the housing to another.

The thickness of the lip base is linearly related to the frequency of lip reed oscillation. This effect completely overrules the frequency-lowering effect of the increased mass of thicker lips.

The use of silicone rubber lip reeds for alternative voice production may be limited by their chaotic behavior. Ultimately all lips demonstrate chaotic behavior, but only for thin lip configurations this occurs well within the regular airflow range for TE shunt valve speech. Chaotic behavior deteriorates the harmonics-to-noise ratio, and would therefore compromise the naturalness and intelligibility of the artificial voice.

High-speed visualization of lip reed motion provides both general insight and details of the underlying lip reed behavior essential for optimization of this artificial voice source.

In addition to previous reports that concerned only female laryngectomees, this study demonstrated that also for male laryngectomees a lip configuration may be found that offers a good compromise between the desired f_0 range and favorable sound quality.

ACKNOWLEDGMENTS

This study was supported by grant GGN 55.3712 from the Dutch Technology Foundation STW. The authors wish to thank C. Klok and D. Koops (Dept. of Clinical Physics and Informatics, VU Medical Center, Amsterdam, The Netherlands) for the computer-aided engineering of the silicone rubber lips and for preparing the perspex TE-model. We acknowledge J. Reinders (Reinka lm-en Export BV, Breda, The Netherlands) for giving us the use of the Weinberger high-speed digital imaging system, and A. Hirschberg (Dept. of Applied Physics, University of Technology, Eindhoven, The Netherlands) for the helpful discussions on aero-acoustics in general and his valuable comments on the initial version of this paper.

[1] E. D. Blom and R. C. Hamaker, Tracheoesophageal voice restoration following total laryngectomy, in *Cancer of the Head and Neck*, edited by E. N. Meyers and J. Suen (WB Saunders, Philadelphia, 1996), pp. 839–852.
[2] M. I. Singer and E. D. Blom, ''An endoscopic technique for restoration of voice after laryngectomy,'' Ann. Otol. Rhinol. Laryngol. **89**, 529–533 (1980).
[3] H. F. Mahieu, H. K. Schutte, and A. A. Annyas, Intelligibility, vocal intensity, and long-term average spectra of Groningen button-oesophageal speech, in *Speech Restoration via Voice Prostheses*, edited by I. F. Hermann (Springer-Verlag, Berlin, Heidelberg, New York, 1986), pp. 139–147.
[4] H. K. Schutte, E. F. M. Bors, G. H. A. de Boer, G. L. J. Nieboer, and A. A. Annyas, Evaluation of speech with and without a 'Groningen type' voice button, in *Speech Restoration via Voice Prostheses*, edited by I. F. Hermann (Springer-Verlag, Berlin, Heidelberg, New York, 1986), pp. 135–138.
[5] L. Max, W. Steurs, and W. de Bruyn, ''Vocal capacities in esophageal and tracheoesophageal speakers,'' Laryngoscope **106**, 93–96 (1996).
[6] A. D. Cheesman, J. Knight, J. Mclvor, and A. Perry, ''Tracheo-oesophageal puncture speech. An assessment technique for failed oesophageal speakers,'' J. Laryngol. Otol. **100**, 191–199 (1986).
[7] H. F. Mahieu, A. A. Annyas, H. K. Schutte, and E. J. van der Jagt, ''Pharyngoesophageal myotomy for vocal rehabilitation of laryngectomees,'' Laryngoscope **97**, 451–457 (1987).
[8] M. I. P. Krook, ''Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency analysis,'' Folia Phoniatr. **40**, 82–90 (1988).
[9] M. D. Trudeau and Y. Qi, ''Acoustic characteristics of female tracheoesophageal speech,'' J. Speech Hear. Dis. **55**, 244–250 (1990).
[10] M. P. de Vries, A. van der Plaats, M. van der Torn, H. F. Mahieu, H. K. Schutte, and G. J. Verkerke, ''Design and *in vitro* testing of a voice-producing element for laryngectomized patients,'' Int. J. Artif. Organs **23**, 462–472 (2000).
[11] M. van der Torn, M. P. de Vries, J. M. Festen, I. M. Verdonck-de Leeuw, and H. F. Mahieu, ''Female-pitched sound-producing voice prostheses-initial experimental and clinical results,'' Eur. Arch. Oto-Rhino-Laryngol. (in press).
[12] D. P. Shedd, V. Bakamjian, K. Sako, M. Mann, S. Barba, and N. Schaaf,

''Reed-fistula method of speech rehabilitation after laryngectomy,'' Am. J. Surg. **124**, 510–514 (1972).
[13] R. Hagen, K. Berning, M. Korn, and F. Schon, ''Voice prostheses with sound-producing metal reed element—an experimental study and initial clinical results,'' Laryngorhinootologie **77**, 312–321 (1998).
[14] N. H. Fletcher, ''Autonomous vibration of simple pressure-controlled valves in gas flows,'' J. Acoust. Soc. Am. **93, Pt. 1**, 2172–2180 (1993).
[15] F.-C. Chen and G. Weinreich, ''Nature of the lip reed,'' J. Acoust. Soc. Am. **99**, 1227–1233 (1996).
[16] M. van der Torn, I. M. Verdonck-de Leeuw, J. M. Festen, M. P. de Vries, and H. F. Mahieu, ''Alternative voice after laryngectomy using a sound-producing voice prosthesis,'' Laryngoscope **111**, 336–346 (2001).
[17] M. P. de Vries, M. C. Hamburg, A. E. P. Veldman, H. K. Schutte, and G. J. Verkerke, Numerical simulation of a voice-producing element based on Navier Stokes equitations and the Finite Element Method, in A new voice for the voiceless, Ph.D. dissertation by M. P. de Vries, University of Groningen, Groningen, The Netherlands, 2000, Chap. 4, pp. 57–73.
[18] D. C. Copley and W. J. Strong, ''A stroboscopic study of lip vibrations in a trombone,'' J. Acoust. Soc. Am. **99**, 1219–1226 (1996).
[19] A. E. Rosenberg, ''Effect of glottal pulse shape on the quality of natural vowels,'' J. Acoust. Soc. Am. **49**, 583–990 (1971).
[20] M. R. Schroeder, ''Period histogram and product spectrum: new methods for fundamental frequency measurement,'' J. Acoust. Soc. Am. **43**, 829–834 (1968).
[21] G. de Krom, ''A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals,'' J. Speech Hear. Res. **36**, 254–266 (1993).
[22] J. L. Flanagan, *Speech Analysis—Synthesis and Perception*, 2nd ed. (Springer-Verlag, Berlin, Heidelberg, New York, 1972), p. 154.
[23] J. M. Festen, J. H. M. van Beek, H. F. Mahieu, and A. J. Parker, Acoustic analysis of tracheoesophageal voice, in *Surgery and Prosthetic Voice Restoration after Total and Subtotal Laryngectomy*, edited by J. Algaba (Elsevier Science BV, Amsterdam, 1996), pp. 171–175.
[24] R. D. Ayers, New perspectives on the brass instruments, in *Proceedings of the International Symposium on Musical Acoustics*, edited by D. Keefe, T. Rossing, and C. Schmid (Acoust. Soc. Am., Leavenworth, 1998), pp. 132–134.
[25] A. van der Plaats, M. P. de Vries, F. J. M. van der Eerden, H. K. Schutte, and G. J. Verkerke, In-vitro measurements of a voice-producing element under physiologic acoustic conditions, in A new voice for the voiceless, Ph.D. dissertation by M. P. de Vries, University of Groningen, Groningen, The Netherlands, 2000, Chap. 5, pp. 75–93.
[26] B. H. Story, I. R. Titze, and E. A. Hoffman, ''Vocal tract area functions from magnetic resonance imaging,'' J. Acoust. Soc. Am. **100**, 537–554 (1996).
[27] K. Ishizaka, M. Matsudaira, and T. Kaneko, ''Input acoustic-impedance measurement of the subglottal system,'' J. Acoust. Soc. Am. **60**, 190–197 (1976).
[28] X. Pelorson, A. Hirschberg, A. P. J. Wijnands, and H. Bailliet, ''Description of the flow through in-vitro models of the glottis during phonation,'' Acta Acust. (Beijing) **3**, 191–202 (1995).
[29] G. Fairbanks, *Voice and Articulation Drillbook* (Harper and Row, New York, 1960).
[30] I. R. Titze, ''On the relation between subglottal pressure and fundamental frequency in phonation,'' J. Acoust. Soc. Am. **85**, 901–906 (1989).
[31] Y. Aurégan and C. Depollier, ''Snoring: Linear stability analysis and in-vitro experiments,'' J. Sound Vib. **188**, 39–54 (1995).
[32] A. Hirschberg, R. W. A. van de Laar, J. P. Marrou-Maurières, A. P. J. Wijnands, J. H. Dane, S. G. Kruijswijk, and A. J. M. Houtsma, ''A quasi-stationary model of air flow in the reed channel of single-reed woodwind instruments,'' Acustica **70**, 146–154 (1990).
[33] G. R. Plitnik, ''Vibration characteristics of pipe organ reed tongues and the effect of the shallot, resonator and reed curvature,'' J. Acoust. Soc. Am. **107**, 3460–3473 (2000).
[34] J. M. Heaton, D. Sanderson, I. R. Dunsmore, and A. J. Parker, ''*In vivo* measurements of indwelling tracheo-oesophageal prostheses in alaryngeal speech,'' Clin. Otolaryngol. **21**, 292–296 (1996).
[35] E. B. Holmberg, R. E. Hillman, and J. S. Perkell, ''Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice,'' J. Acoust. Soc. Am. **84**, 511–529 (1988).
[36] R. F. Coleman, J. H. Mabis, and J. K. Hinson, ''Fundamental frequency-sound pressure level profiles of adult male and female voices,'' J. Speech Hear. Res. **20**, 197–204 (1977).
[37] H. Hollien, D. Dew, and P. Phillips, ''Phonational frequency ranges of adults,'' J. Speech Hear. Res. **14**, 755–760 (1971).

Measuring vocal quality with speech synthesis

Bruce R. Gerratt[a] and Jody Kreiman[b]
Division of Head and Neck Surgery, UCLA School of Medicine, 31-24 Rehab Center, Los Angeles, California 90095-1794

(Received 7 December 2000; accepted for publication 13 August 2001)

Much previous research has demonstrated that listeners do not agree well when using traditional rating scales to measure pathological voice quality. Although these findings may indicate that listeners are inherently unable to agree in their perception of such complex auditory stimuli, another explanation implicates the particular measurement method—rating scale judgments—as the culprit. An alternative method of assessing quality—listener-mediated analysis-synthesis—was devised to assess this possibility. In this new approach, listeners explicitly compare synthetic and natural voice samples, and adjust speech synthesizer parameters to create auditory matches to voice stimuli. This method is designed to replace unstable internal standards for qualities like breathiness and roughness with externally presented stimuli, to overcome major hypothetical sources of disagreement in rating scale judgments. In a preliminary test of the reliability of this method, listeners were asked to adjust the signal-to-noise ratio for 12 synthetic pathological voices so that the resulting stimuli matched the natural target voices as well as possible For comparison to the synthesis judgments, listeners also judged the noisiness of the natural stimuli in a separate task using a traditional visual-analog rating scale. For 9 of the 12 voices, agreement among listeners was significantly (and substantially) greater for the synthesis task than for the rating scale task. Response variances for the two tasks did not differ for the remaining three voices. However, a second experiment showed that the synthesis settings that listeners selected for these three voices were within a difference limen, and therefore observed differences were perceptually insignificant. These results indicate that listeners can in fact agree in their perceptual assessments of voice quality, and that analysis-synthesis can measure perception reliably.
[DOI: 10.1121/1.1409969]

PACS numbers: 43.71.Bp, 43.71.Gv [KRK]

I. INTRODUCTION

The study of voice quality has always been challenging. Difficulties arise at the definitional level (e.g., Jensen, 1965; Sundberg, 1987) and compound at every step of investigation. The links between perceived quality and the underlying vocal physiology or acoustics are not well understood, so the importance of perceptual voice features cannot be established by reference to objective measures of phonatory function or the acoustic waveform. The inherently multidimensional nature of voice quality increases the complexity of measurement and interpretation of results. Further, no satisfactory basis exists for defining an *a priori* structure for the perceptual space, and the nature of a set of features that might specify voice quality remains unknown. Because of this, instrumental measures of voice cannot be validated in a straightforward manner by their relationship to independently validated perceptual constructs. Finally, the appropriate method for measuring what listeners hear remains an unresolved issue. Research on voice quality perception has used rating scale techniques (including visual analog scales, equal-appearing interval scales, and direct magnitude estimation) almost exclusively. However, listeners often disagree in their assessment of overall voice quality and in their use of scales for individual voice qualities like roughness and breathiness (Gelfer, 1988; Kreiman and Gerratt, 1998; Orlikoff, 1999; Wuyts *et al.*, 1999; but cf. Hillenbrand *et al.*, 1994; Hillenbrand and Houde, 1996). [See, e.g., Poulton (1979) and Gescheider (1997) for review of similar issues regarding judgments of the loudness of pure tones.] Difficulty isolating single perceptual dimensions of complex, multidimensional stimuli appears to be a significant source of listener disagreement (Kreiman and Gerratt, 2000; Kreiman *et al.*, 1994). Such listener disagreement reduces confidence in the validity and utility of existing perceptual measures of voice, further undermining attempts to define, describe, or quantify vocal quality.

Finding valid and reliable alternatives to traditional voice quality scaling methods requires hypotheses about the sources of listener disagreements, so that techniques can be developed to control such variability. Previous studies of pathological voices (Gerratt *et al.*, 1993; Kreiman and Gerratt, 2000) suggest that traditional perceptual scaling methods are effectively matching tasks, where external stimuli (the voices) are compared to stored mental representations that serve as internal standards for the various rating scales. These idiosyncratic, internal standards appear to vary with listeners' previous experience with voices (Kreiman *et al.*, 1990; Verdonck-de Leeuw, 1998) and with the context in which a judgment is made (Gerratt *et al.*, 1993; cf. Gescheider and Hughson, 1991), and may vary substantially across listeners as well as within a given listener (Gerratt

[a]Electronic mail: bgerratt@ucla.edu
[b]Electronic mail: jkreiman@ucla.edu

et al., 1993; Kreiman *et al.*, 1993). In addition, severity of vocal deviation, difficulty isolating individual dimensions in complex perceptual contexts, and factors like lapses in attention can also influence perceptual measures of voice (de Krom, 1994; Kreiman and Gerratt, 2000). These factors (and possibly others) presumably all add uncontrolled variability to scalar ratings of vocal quality, and contribute to listener disagreement.

This article describes a preliminary test of an alternative approach to the assessment of voice quality that is designed to control these sources of measurement error. In this method, listeners vary speech synthesis parameters to create an acceptable auditory match to a natural voice stimulus. When a listener chooses a best match to a test stimulus, the synthesis settings parametrically represent the listener's perception of voice quality. Because listeners directly compare each synthetic token they create to the target natural voice, they need not refer to internal standards for particular voice qualities. Further, listeners can manipulate acoustic parameters and hear the result of their manipulations immediately. We hypothesize that this process will help listeners focus attention on individual acoustic dimensions, reducing the perceptual complexity of the assessment task and presumably the associated response variability. In theory, then, the analysis-synthesis method should improve agreement among listeners in their assessments of voice quality relative to traditional rating scale techniques, because it controls the major sources of variance in quality judgments.

Similar logic motivated a previous attempt to reduce variability in ratings of voice quality through the use of fixed external reference stimuli (perceptual ''anchors'') (Gerratt *et al.*, 1993). Listeners in that study assessed vocal roughness of synthetic stimuli using a five-point scale in which a synthetic voice sample exemplified each scale value. Because listeners compared test stimuli to these ''anchor'' stimuli, we hypothesized that they would agree better in their ratings than they would when referring only to their internal criteria for different levels of roughness in a traditional rating scale task. In fact, listener agreement did increase significantly when test stimuli were identical to or immediately adjacent to the anchor stimuli, but agreement decreased sharply for stimuli that fell further from the anchors in the series [cf. Wedell *et al.* (1990), who found similar effects in ratings of the severity of psychiatric symptoms]. Presumably, the further apart the test and anchor stimuli were acoustically, the more listeners relied on their internal quality standards, resulting in lower interrater agreement. Thus, while the anchored task reduced overall variation in listener ratings of voice quality, the limited scale resolution proved a significant source of measurement error.

The analysis-synthesis technique described in this report is designed to correct this limitation. This task provides the same theoretical advantages as the anchored protocol, in that listeners explicitly match reference and test stimuli. However, the analysis-synthesis task provides much finer scale resolution, allowing listeners to create a very close match to the perceived quality of the test voice. Anchored protocols also require the experimenter to choose the specific unidimensional quality scales along which voices are to be rated (e.g., breathiness, roughness, etc.). Such scales have resisted empirical validation, and the nature of the ''correct'' set of scalar features for voice quality remains unknown. In contrast, analysis-synthesis techniques permit assessment of quality without the need to postulate discrete perceptual scales for particular aspects of quality, because synthesis parameters combine to model overall vocal quality. These theoretical advantages should provide better measurement reliability and validity for a synthesis rating task relative to traditional scalar voice rating techniques. The experiment reported below provides a preliminary test of this hypothesis.

II. EXPERIMENT 1

In this preliminary test of the analysis-synthesis method of evaluating voice quality, listeners were asked to manipulate a single acoustic parameter. Although voice quality is multidimensional, restricting the task in this way simplifies comparing the reliability of the synthesis task to that of traditional rating scale measures of voice, and presents the simplest case to test the potential efficacy of this technique. The signal-to-noise ratio (SNR) was selected because it has been historically important in describing voice quality (e.g., Hirano, 1988; Michaelis *et al.*, 1998; Yumoto *et al.*, 1982), and because it can be manipulated with a single parameter (unlike the periodic component of the source, for example, which must be specified with several parameters as described below). In a separate experiment, listeners also evaluated the ''noisiness'' of the stimulus voices using a traditional visual analog rating scale. To the extent that the synthesis protocol controls sources of interrater variability, listeners should agree better in their choice of synthesis settings than they do in the analogous rating scale task.

A. Method

1. Voice samples

Twelve samples of the vowel /a/ were selected from a set recorded from patients in a clinical setting. Pathological voice samples were studied for two reasons. First, accurate and reliable measurement of voice quality is of particular clinical importance, because patients and clinicians are often concerned about the nature and extent of vocal deviation. Further, pathological voices encompass a wide range of the human larynx's phonatory potential, and samples with pathology have a greater range of vocal quality variables than do samples of normal phonation. Speakers ranged in age from 26 to 73 years (mean = 46.3 years), and represented a variety of primary diagnoses, including essential vocal tremor, vocal fold mass lesions, vocal fold paralysis, adductory spasmodic dysphonia, reflux laryngitis, glottal incompetence, and laryngeal web. They ranged from mildly to severely dysphonic. Both male ($n=4$) and female speakers ($n=8$) were included.

Voice signals were transduced with a 1-in. Bruel and Kjaer condenser microphone held a constant 5 cm off axis. They were then low-pass filtered at 8 kHz and directly digitized at 20 kHz. A 1-s segment was excerpted from the middle of these productions, antialias filtered, and downsampled to 10 kHz for further analysis.

2. Analysis and synthesis

A synthetic copy of each voice was created by an experienced operator using a custom formant synthesizer implemented in MATLAB (MathWorks, Natick, MA).[1] The synthesizer operates in near-real time, and allows users to manipulate the fundamental frequency ($f0$), the shape of the glottal volume velocity derivative, the spectrum of the inharmonic component of the voice source (the noise spectrum), the overall signal-to-noise ratio, formant frequencies and bandwidths, and the rate, extent, and regularity of frequency modulation (tremor).

The goal of synthesis at this stage of the investigation was to create the best possible copy of each target voice, so that listener performance later in the experiment could be attributed to task factors, rather than to limitations of the stimuli. Initial parameter settings for synthesis were derived from acoustic analyses of the voices as follows. Formant frequencies and bandwidths were estimated using autocorrelation linear predictive coding (LPC) analysis with a window of 25.6 ms (increased to 51.2 ms when stimulus $f0$ was near or below 100 Hz). The glottal volume velocity derivative was estimated by inverse filtering a single glottal pulse from the microphone recordings, using the method described by Javkin *et al.* (1987). The frequency of this cycle served as the initial estimate of mean $f0$.

The output of the inverse filter was least-squares fit with a Liljencrants–Fant (LF) source model (Fant *et al.*, 1985), and the parameters of the best-fitting LF model were used to specify the harmonic component of the voice source in the synthesizer. The traditional LF model was slightly modified to improve the fit of the return phase, and so that the beginning of the closed phase was explicitly specified with a parameter. The ''equal area constraint,'' which requires that areas under the positive and negative curves in the flow derivative be the same, was also abandoned. When modeling some voices with this constraint in place, the return phases did not consistently return to zero. This introduced high-frequency artifacts when the next pulse began at 0 (see Epstein *et al.*, 1999, for further details).

Slow variations in $f0$ (vocal tremors) were modeled by modulating the nominal $f0$ in one of two patterns: a sine wave, or irregular modulation. Rate and extent of frequency modulation in both cases were estimated from $f0$ tracks of the natural voices (see Kreiman *et al.*, 2001, for details).[2]

To synthesize noise, a cepstral-domain comb filter similar to that described by de Krom (1993) first removed the harmonic part of the signal, leaving an estimate of the inharmonic component of the voice. This residual was then inverse filtered to remove the effects of vocal tract resonances, leaving the inharmonic part of the source. Next, this noise spectrum was fitted with a 25-segment piece-wise linear approximation. Finally, a 100-tap finite impulse response filter was synthesized for the fitted noise spectrum, and a spectrally shaped noise time series was created by passing white noise through this filter. The effects of jitter and shimmer were not modeled separately from overall spectral noise.

The synthesizer sampling rate was fixed at 10 kHz. To overcome quantization limits on modeling $f0$, the source time series was synthesized pulse by pulse using an interpolation algorithm. Within each pulse, samples were interpolated at exact sample instants as follows. A plot of $f0$ versus time was generated for the duration of the 1-s token to be synthesized, taking into account any modeled vocal tremor. Source pulses with frequencies dictated by the $f0$ versus time plot were calculated, and then concatenated to produce a synthetic time series. To reduce phase error, the absolute beginning and ending times of each pulse were tracked and used in the interpolation of succeeding pulses. At the beginning instant of each new pulse (which could occur at any time, including between samples), the $f0$ curve generated above was interpolated to find $f0$ for this LF pulse. Given that $f0$, the LF pulse was stretched or compressed to obtain the appropriate period, and sample points were calculated accordingly. The overall effect is equivalent to digitizing an analog pulse train with pulses of the exact desired frequencies at the fixed 10-kHz sample rate.

The LF pulse train was added to the noise time series previously described to create a complete glottal source waveform. The ratio of noise to LF energy was initially set to match the value calculated from the original voice sample. The complete synthesized source was filtered through the vocal tract model (estimated through LPC analysis, as described earlier) to generate a preliminary version of the synthetic voice. Finally, the operator adjusted all synthesis parameters to achieve the best possible perceptual match to the original voice. Note that all synthesis parameters, including noise, were optimized simultaneously. Again, the goal at this stage in the experiment was to create the best possible overall match to each target voice. In the judgment of listeners in a previous experiment (Kreiman *et al.*, 2001), all 12 synthetic stimuli provided excellent matches to the original voices. In particular, spectral noise integrated into the stimuli, resulting in very natural-sounding synthetic voices.

3. Listeners

Ten expert listeners (five otolaryngologists, three speech-language pathologists, and two phoneticians, including both authors, one of whom created the stimuli) participated in this experiment. All had extensive experience evaluating and/or treating voice disorders, and all reported normal hearing.

4. Experimental synthesis task

Prior to the experiment the signal-to-noise ratio for each synthetic stimulus was set to 50 dB. This produced synthetic voices that were free of noise. Listeners were then asked to change the overall signal-to-noise ratio until the synthetic token perceptually matched the natural target token. Listeners made their adjustments by moving a sliding cursor with a mouse along a 100-mm scale displayed on a computer monitor.[3] Left and right endpoints of this scale corresponded to signal-to-noise ratios of 50 dB (noise free) and 0 dB (high level of noise). These limits spanned the range of noise levels found in pilot studies of 70 pathological voices. Adjustments could be made in steps of 0.05 dB. Changes in the position of the cursor increased or decreased the overall

TABLE I. Rating variances for the two perceptual tasks.

Voice	Rating variance: Synthesis task	Rating variance: Visual analog rating task	$F(9,9)$
1	6.27	234.54	37.41, $p<0.05$
2	35.13	806.27	22.95, $p<0.05$
3	10.27	407.78	39.71, $p<0.05$
4	17.34	131.96	7.61, $p<0.05$
5	485.88	144.10	3.37, n.s.
6	13.97	86.46	6.19, $p<0.05$
7	33.69	157.51	4.68, $p<0.05$
8	30.97	356.27	11.50, $p<0.05$
9	433.49	317.73	1.36, n.s.
10	3.58	241.96	67.59, $p<0.05$
11	295.52	402.23	1.36, n.s.
12	11.42	554.18	48.52, $p<0.05$

signal-to-noise ratio, but did not change the shape of the noise spectrum, which remained fixed for each voice throughout the experiment. Within each trial and across all listeners, other synthesis parameters were held constant at values that produced good perceptual matches to the original voices (as determined by listeners in Kreiman *et al.*, 2001). This provided a constant perceptual frame for listeners' noise manipulations.

The experiment took place in a double-walled sound booth. Listeners heard the stimuli in free field over good quality loudspeakers, at a comfortable listening level. Voices were presented in a different random order to each listener. Listeners were able to play the synthetic token and natural target stimulus as often as necessary, and could make as many adjustments to the noise level as necessary to achieve a satisfactory match during their noise adjustments. Testing took approximately 40 min.

5. Rating scale task

Ten expert listeners (eight of whom also participated in the synthesis experiment) judged the perceived noisiness of the 12 original, natural, pathologic stimuli using a traditional 100-mm visual-analog rating scale whose two ends were labeled ''no noise'' and ''extremely noisy.'' The scale was displayed on a computer monitor, and was the same length, color, and shape as that used in the synthesis task. Listeners made their judgments by moving a sliding cursor with a mouse, as above. Listeners were able to replay the voices as often as necessary prior to making their responses. Stimuli were rerandomized for each listener. When listeners participated in both tasks, order of presentation was also randomized, and tasks were separated by at least 1 week.

B. Results

Because the signal-to-noise ratio ranged from 50 to 0, listener responses in the synthesis task were multiplied by 2 prior to statistical analysis. Thus, possible response values for both tasks ranged from 0 to 100. Table I shows the response variances for each voice, for the synthesis and visual analog rating tasks. For 9 of the 12 voices, variance in synthesis settings was significantly (and substantially) less than that for the visual analog ratings. Rating variances in the two tasks did not differ significantly for the remaining three voices.

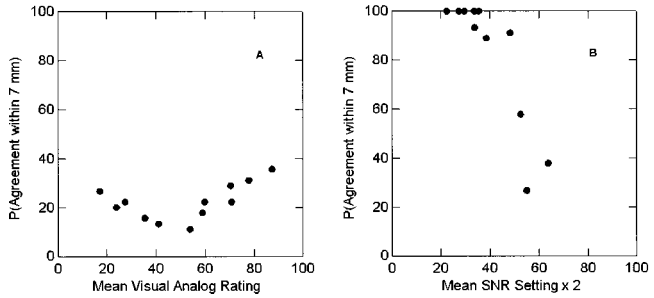


FIG. 1. For each voice, the likelihood of two raters agreeing within 7 mm in their voice assessments versus the mean rating for that voice. (a) The visual analog scale task. (b) The analysis-synthesis task. One case is hidden by overlapping points.

To compare patterns of listener agreement, we calculated the likelihood that two listeners' ratings would agree within 7 units on the 100 unit scales [corresponding to approximately 0.5 scale value on a 7 point equal-appearing interval scale (Kreiman and Gerratt, 1998)]. These probabilities are plotted against the mean rating for each voice in Fig. 1. Patterns of agreement for the visual analog task showed near zero agreement between pairs of listeners for voices whose mean rating was near the middle of the scale, with slightly better agreement for voices with mean ratings at the ends of the scale [Fig. 1(a)], similar to patterns observed for other voice quality scales (Kreiman and Gerratt, 1998). In contrast, the probability of agreement between pairs of listeners was at or near ceiling for nine stimuli in the synthesis task, and decreased for the three voices that received the highest mean SNR settings [Fig. 1(b)]. The cause of this apparent decrease in interrater reliability for these voices is examined in experiment 2, reported later in this work.

C. Discussion

These results indicate that the synthesis protocol does reduce variability in measures of perceived voice quality relative to a traditional visual analog scale task. Listeners agreed perfectly or nearly perfectly in their perceptual evaluations of pathological voice stimuli when they were given a suitable tool for reporting their judgments. In particular, listeners agreed at or near ceiling levels for the six stimuli in the midrange of mean SNR responses. To our knowledge, agreement in the midrange never occurs for traditional rating protocols (Kreiman and Gerratt, 1998). Because listener agreement improved markedly when variability due to task-related factors was controlled, listeners' disagreements in traditional rating scale tasks can be attributed to task-related difficulties, rather than to differences between listeners in perceptual processes or to the inherently subjective nature of voice perception (Weismer and Liss, 1991).

III. EXPERIMENT 2

Some of the remaining interrater variability in the synthesis task (noted particularly in 3 of the 12 voices) may be due to the lack of perceptual calibration of the signal-to-

noise ratio scale. That is, differences among responses along this scale may not exceed a difference limen, and thus may not represent disagreement at all. Experiment 2 examined listeners' ability to distinguish differences in levels of noise for the stimuli in experiment 1, to determine if the apparent response variability in experiment 1 is perceptually significant.

A. Method

1. Stimuli

Twelve series of stimuli were synthesized, one series for each of the 12 voices in experiment 1, using the custom synthesizer previously described. Stimuli within each series varied only in noise levels. The minimum value for each series corresponded to the lowest level of noise selected for that voice by the listeners in experiment 1, and the maximum value corresponded to the highest level of noise selected for that voice. Four equal noise steps were created between these levels, for a total of six versions of each original voice. All other synthesis parameters were held constant within a voice "family."

Each stimulus was 1 s in duration. Stimuli were scaled for equal peak amplitude and multiplied by 30-ms ramps to eliminate click artifacts.

2. Listeners

Twenty listeners participated in this experiment. All reported normal hearing. Both expert ($n=5$) and nonexpert listeners (UCLA students; $n=15$) participated, because the task required a same/different judgment, and experts have been shown to differ from naïve listeners only for scalar ratings of individual qualities like those used in experiment 1 (e.g., Kreiman *et al.*, 1990).

3. Procedure

Listeners heard all possible pairs of the six different synthetic stimuli in each series, along with an equal number of pairs where the stimuli were identical, for a total of 360 trials/listener. Stimuli within a pair were separated by 500 ms. Stimulus pairs were rerandomized for each listener.

For each pair of voices, listeners judged whether the two stimuli were the same or different. They also rated their confidence in their response, on a five-point scale ranging from "wild guess" to "positive." These confidence ratings were used to derive receiver operating characteristics (ROCs), as described in the next section.

Testing took place in a double-walled sound booth. Stimuli were presented in free field over good quality loudspeakers. Listeners were able to replay each pair of voices as often as necessary before responding. Testing lasted approximately 45 min.

B. Results and discussion

Same/different responses were combined with listeners' confidence ratings to create a ten-point scale ranging from "certain that voices are different" to "certain that voices are the same." Receiver operating characteristics (ROCs) were constructed from these recoded data by plotting the probability of an incorrect "same" response on the x axis versus the probability of a correct "same" response on the y axis, for each level of the ten-point scale. The area under the resulting curve ranges from 0.5 (chance) to 1.0 (perfect performance) and is a measure of listeners' ability to discriminate among the different stimuli that is independent of biases in favor of "same" or "different" responses (see, e.g., Egan, 1975, or Swets and Pickett, 1982, for review).

Recall that for each six-member series of stimuli, listeners heard pairs of voices that differed in noise levels by one to five steps. To assess listeners' ability to hear larger or smaller differences in noise level, separate ROCs were constructed for stimuli differing by one step, two steps, three steps, or four to five steps, for a total of four ROCs for each of the 12 series. The 99% confidence interval around each ROC (as calculated by SPSS; SPSS Inc., Chicago, IL) was used to determine whether listeners were able to discriminate stimuli beyond chance levels. When the 99% confidence interval for the area under a given ROC included the chance value of 0.5, we concluded that differences of that size in noise steps were not perceptually significant.

No difference between expert and nonexpert listeners was found in overall discrimination accuracy [$F(1,18)=0.03$, n.s.]. To interpret the results of experiment 1, we calculated the differences between the signal-to-noise ratio levels selected by all possible pairs of raters, and evaluated the significance of each difference with respect to the ROCs derived earlier. For example, if a pair of raters differed in their chosen noise settings by between two and three noise steps, and stimuli differing by three steps were not discriminable in the present experiment, those two responses were considered to be perceptually equivalent. These analyses indicated that 72.5% of the responses produced by the listeners in experiment 1 were in fact perceptually identical (area under the ROC$\leqslant$0.5). An additional 14.2% of responses were significantly but poorly discriminable (0.6$<$area under the ROC $<$0.7; hit rate=86.6%; false alarm rate=58.0%, where a hit was defined as a correct "same" response). Only 3 listener-selected SNR values out of the 120 in experiment 1 (2.5%) resulted in stimuli that were consistently discriminable from the other members of their voice "families" (area under the ROC$\geqslant$0.9). These three signal-to-noise ratio responses were generated by different listeners, and were intended to match different target voices, supporting the view that they represent random rather than systematic errors (perhaps due to lapses in attention).

Recall that for three voices in experiment 1 (Nos. 5, 9, and 11; Table I), ratings from the synthesis and visual analog tasks did not differ in variance, and the likelihood of close agreement between pairs of listeners for these three voices in the synthesis task was relatively poor. ROC analyses for these voices indicated that, for voices 5 and 9, all tokens were perceptually identical (area under the ROC$\leqslant$0.5), despite the differences in signal-to-noise ratios. For voice 11, nine tokens were indiscriminable, while the tenth was only poorly discriminable from the others (area under the ROC $=0.63$). Thus, the majority of apparent listener disagreements in the analysis-synthesis task in fact resulted from the

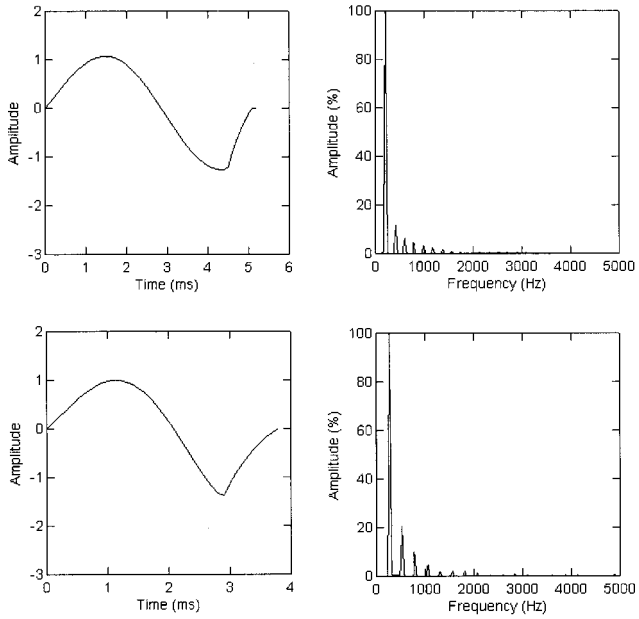


FIG. 2. Glottal volume velocity derivatives and associated linear source spectra for two voices for which listeners agreed well in their choice of signal-to-noise ratios.

fact that most of the responses were within a difference limen of each other.

Examination of estimated source spectra provides a preliminary explanation for the differences in listeners' sensitivity to changes in noise levels across stimuli. Figure 2 shows representative glottal volume velocity derivatives and the associated linear fast Fourier transform spectra for the synthetic stimuli about which listeners agreed precisely in their selection of noise levels. Figure 3 shows the volume velocity derivatives and source spectra for voices 5, 9, and 11, for which listeners varied in their chosen signal-to-noise ratio. The voices about which listeners agreed well in their noise settings all had sources with very little energy above the first or second harmonic. In contrast, voices 5, 9, and 11 all had sources with significant amounts of high-frequency harmonic energy, in addition to the excitation provided by the aperiodic component of the source [cf. Cranen and Schroeter (1995), who describe the different effects of posterior versus medial glottal gaps on source spectral slopes]. These two groups of stimuli differed significantly in values of the LF composite parameter RA (Fant and Lin, 1988), which measures the amount of harmonic source energy in the higher frequencies [$F(1,10)=75.35$, $p<0.05$]. Although any interpretation must be tentative due to the limited amount of data, small differences in noise levels are apparently more difficult to discriminate in the presence of harmonic energy in the higher frequencies, presumably due to masking effects. Unlike rating scale tasks, the analysis-synthesis protocol could be used in a straightforward manner to test this hypothesis (and others like it), and to derive the precise relationship between listener sensitivity and the characteristics of the harmonic and inharmonic components of the source.

IV. GENERAL DISCUSSION

These results indicate that listeners do in fact agree in their perceptual assessments of pathological voice quality,

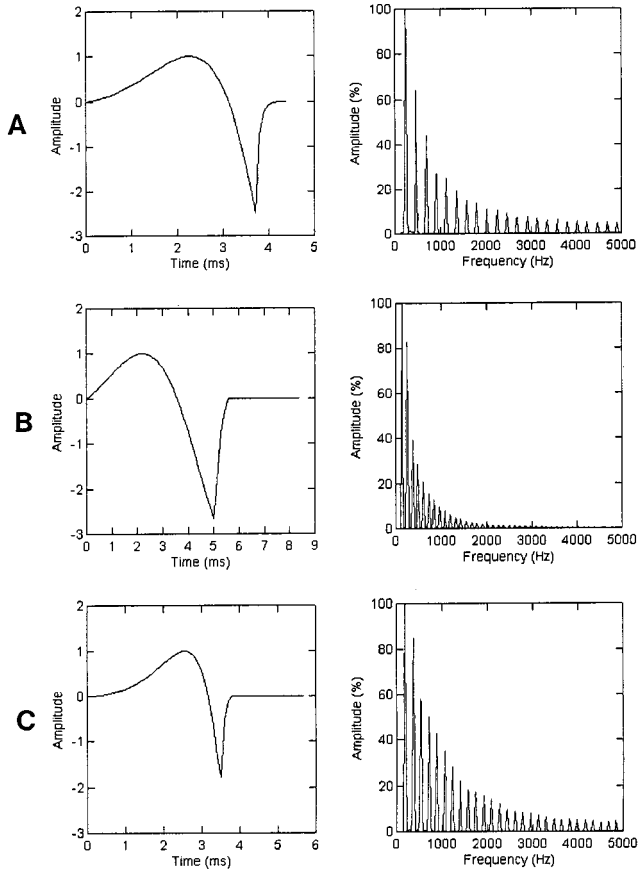


FIG. 3. Glottal volume velocity derivatives and associated linear source spectra for the three voices for which listeners gave varying responses when asked to match the signal-to-noise ratio. Voice numbers correspond to Table I: (a) voice 5, (b) voice 9, and (c) voice 11.

and that tools can be devised to measure perception reliably. Many previous studies have demonstrated disagreement among listeners in their use of traditional unidimensional rating scales for voice quality. However, the use of an analysis-synthesis protocol greatly simplifies the task of assessing quality. It does not require listeners to refer to varying internal standards for individual qualities, and facilitates attending to single acoustic dimensions in complex, varying contexts. Thus, this task appears to control several major sources of variability in listeners' responses, providing perceptual measures that are relatively free of the confounding influences of cognitive and task-related variables.

A reliable, valid protocol for quantifying voice quality perception would have many applications. Researchers and clinicians alike have long sought objective measures of voice quality, for example to examine prosodic or linguistic changes in voice quality, to model differences between speakers in individual voice quality, or to track improvement within a single speaker with treatment. One traditional research approach has been to gather scalar ratings for individual voice qualities and calculate correlations between average ratings and acoustic or other instrumental measures of the rated voices. Obviously, this method cannot establish a direct link between the perception of a specific voice and the acoustic element that gave rise to the perception, due to the limitations of correlational approaches and to the fact that listener disagreements undermine the interpretation of aver-

age ratings. The analysis-synthesis method allows listeners to record their perception of voice quality parametrically and objectively by selecting the level of a set of acoustic attributes. Because this approach uses speech synthesis, signals can be manipulated systematically to demonstrate direct causation between acoustic parameters and perceived quality, overcoming the first of these limitations; and the results presented here suggest that the listening task produces reliable listener responses, removing the second impediment to understanding the complex association between a signal and its perception. Much more research is certainly needed to determine a meaningful, parsimonious set of acoustic parameters that successfully characterizes all possible normal and pathological voice qualities. However, such a set could obviate the need for voice quality labels, allowing researchers and clinicians to replace quality labels with acoustic parameters whose levels objectively, completely, and validly specify the voice quality of interest.

ACKNOWLEDGMENTS

The research was supported by NIDCD Grant No. DC01797. Thanks are due to Melissa Epstein for creating stimuli and testing listeners, and to Norma Antonanzas and Brian Gabelman for extensive and creative programming support. We are also grateful to Keith Kluender, James Hillenbrand, and two anonymous reviewers for their thoughtful comments on a previous version of this article.

[1]This software is available at http://www.surgery.medsch.ucla.edu/glottalaffairs/software.

[2]A more recent version of this software allows $f0$ tracks to be generated and used instead of computed tremors, if desired.

[3]Exact values of the signal-to-noise ratio could also be typed into a small window next to the sliding cursor, but listeners in general did not use this option.

Cranen, B., and Schroeter, J. (**1995**). ''Modeling a leaky glottis,'' J. Phonetics **23**, 165–177.

de Krom, G. (**1993**). ''A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals,'' J. Speech Hear. Res. **36**, 254–266.

de Krom, G. (**1994**). ''Consistency and reliability of voice quality ratings for different types of speech fragments,'' J. Speech Hear. Res. **37**, 985–1000.

Egan, J. P. (**1975**). *Signal Detection Theory and ROC Analysis* (Academic, New York).

Epstein, M., Gabelman, B., Antoñanzas-Barroso, N., Gerratt, B., and Kreiman, J. (**1999**). ''Source model adequacy for pathological voice synthesis,'' in *Proc. ICPhS99*, pp. 2049–2052.

Fant, G., Liljencrants, J., and Lin, Q. (**1985**). ''A four-parameter model of glottal flow,'' Speech Transmission Lab. Quart. Prog. Status Rep. **4**, 1–13.

Fant, G., and Lin, Q. (**1988**). ''Frequency domain interpretation and derivation of glottal flow parameters,'' STL-QPSR **2-3**, 1–21.

Gelfer, M. P. (**1988**). ''Perceptual attributes of voice: Development and use of rating scales,'' J. Voice **2**, 320–326.

Gerratt, B. R., Kreiman, J., Antoñanzas-Barroso, N., and Berke, G. S. (**1993**). ''Comparing internal and external standards in voice quality judgments,'' J. Speech Hear. Res. **36**, 14–20.

Gescheider, G. A. (**1997**). *Psychophysics: The Fundamentals*, 3rd ed. (Erlbaum, Mahwah, NJ).

Gescheider, G. A., and Hughson, B. A. (**1991**). ''Stimulus context and absolute magnitude estimation: A study of individual differences,'' Percept. Psychophys. **50**, 45–57.

Hillenbrand, J., Cleveland, R., and Erickson, R. (**1994**). ''Acoustic correlates of breathy vocal quality,'' J. Speech Hear. Res. **37**, 769–778.

Hillenbrand, J., and Houde, R. A. (**1996**). ''Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech,'' J. Speech Hear. Res. **39**, 311–321.

Hirano, M., Hibi, S., Yoshida, T., Hirade, Y., Kasuya, H., and Kikuchi, Y. (**1988**). ''Acoustic analysis of pathological voice,'' Acta Oto-Laryngol. **105**, 432–438.

Javkin, H., Antonanzas-Barroso, N., and Maddieson, I. (**1987**). ''Digital inverse filtering for linguistic research,'' J. Speech Hear. Res. **30**, 122–129.

Jensen, P. J. (**1965**). ''Adequacy of terminology for clinical judgment of voice quality deviation,'' Eye Ear Nose Throat Mon. **44**, 77–82.

Kreiman, J., Gabelman, B., and Gerratt, B. R. (**2001**). ''Perceptual and acoustic modeling of vocal tremor,'' submitted for publication.

Kreiman, J., and Gerratt, B. R. (**1998**). ''Validity of rating scale measures of voice quality,'' J. Acoust. Soc. Am. **104**, 1598–1608.

Kreiman, J., and Gerratt, B. R. (**2000**). ''Sources of listener disagreement in voice quality assessment,'' J. Acoust. Soc. Am. **108**, 1867–1879.

Kreiman, J., Gerratt, B. R., and Berke, G. S. (**1994**). ''The multidimensional nature of pathologic vocal quality,'' J. Acoust. Soc. Am. **96**, 1291–1302.

Kreiman, J., Gerratt, B. R., Kempster, G., Erman, A., and Berke, G. S. (**1993**). ''Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research,'' J. Speech Hear. Res. **36**, 21–40.

Kreiman, J., Gerratt, B. R., and Precoda, K. (**1990**). ''Listener experience and perception of voice quality,'' J. Speech Hear. Res. **33**, 103–115.

Michaelis, D., Frohlich, M., and Strube, H. (**1998**). ''Selection and combination of acoustic features for the description of pathologic voices,'' J. Acoust. Soc. Am. **103**, 1628–1639.

Orlikoff, R. O. (**1999**). ''The perceived role of voice perception in clinical practice,'' Phonoscope **2**, 87–106.

Poulton, E. C. (**1979**). ''Models for biases in judging sensory magnitude,'' Psychol. Bull. **86**, 777–803.

Sundberg, J. (**1987**). *The Science of the Singing Voice* (Northern Illinois U.P., De Kalb, IL).

Swets, J. A., and Pickett, R. M. (**1982**). *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory* (Academic, New York).

Verdonck-de Leeuw, I. M. (**1998**). ''Perceptual analysis of voice quality: Trained and naive raters, and self-ratings,'' in *Proceedings of Voicedata98 Symposium on Databases in Voice Quality Research and Education*, edited by G. de Krom (Utrecht Institute of Linguistics OTS, Utrecht), pp. 12–15.

Wedell, D. H., Parducci, A., and Lane, M. (**1990**). ''Reducing the dependence of clinical judgment on the immediate context: Effects of number of categories and type of anchor,'' J. Pers. Soc. Psychol. **58**, 319–329.

Weismer, G., and Liss, J. (**1991**). ''Reductionism is a dead-end in speech research: Perspectives on a new direction,'' in *Dysarthria and Apraxia of Speech: Perspectives on Management*, edited by K. Yorkston, C. Moore, and D. Beukelman (Brookes, Baltimore), pp. 15–27.

Wuyts, F. L., DeBodt, M. S., and Van de Heyning, P. H. (**1999**). ''Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia,'' J. Voice **13**, 508–517.

Yumoto, E., Gould, W. J., and Baer, T. (**1982**). ''Harmonics-to-noise ratio as an index of the degree of hoarseness,'' J. Acoust. Soc. Am. **71**, 1544–1550.

Noisy speech recognition using de-noised multiresolution analysis acoustic features

C. P. Chan, P. C. Ching, and Tan Lee[a)]
Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, People's Republic of China

(Received 20 July 2000; revised 20 March 2001; accepted 4 July 2001)

This paper describes a novel application of multiresolution analysis (MRA) in extracting acoustic features that possess de-noising capability for robust speech recognition. The MRA algorithm is used to construct a mel-scaled wavelet packet filter-bank, from which subband powers are computed as the feature parameters for speech recognition. Wiener filtering is applied to a few selected subbands at some intermediate stages of decomposition. For high-frequency bands, Wiener filters are designed based on a reduced fraction of the estimated noise power, making the consonant features much more prominent and contrastive. The proposed method is evaluated in phone recognition experiments with the TIMIT database. In the presence of stationary white noise at 10-dB SNR, the de-noised MRA features attain a phone recognition rate of 32%. There is a noticeable improvement compared with the accuracy of 29% and 20% attained by the commonly used mel-frequency cepstral coefficients (MFCC) with and without cepstral mean normalization (CMN), respectively. The effectiveness of the MRA features is also verified by the fact that they exhibit smaller distortion from clean speech.
[DOI: 10.1121/1.1398054]

PACS numbers: 43.72.Ar, 43.72.Ne [DOS]

I. INTRODUCTION

The performance of an automatic speech recognition (ASR) system suffers severely from the mismatch of acoustic conditions between training data and test data. Most ASR systems are trained with speech data collected in controlled environments or under noise-free conditions. Substantial degradation of recognition performance is anticipated when these systems are exposed to noise-corrupted input. Numerous methods have been proposed for noisy speech recognition.[1] Many of them are aimed at improving the robustness of the acoustic models. In the hidden Markov model (HMM) framework, this can be done by adapting the model parameters to accommodate noisy speech,[2,3] or alternatively, by decomposing the HMMs into parallel models to deal with signal and noise separately.[4,5] There has also been reported work on the use of robust acoustic features for speech recognition.[6,7] On the other hand, various front-end signal-processing techniques have been investigated with an attempt to remove or subtract noise from the corrupted signal prior to feature extraction. One of the most widely used methods is the so-called cepstral mean normalization (CMN). CMN is a simple but effective technique of compensating acoustic mismatch by performing additive adjustments in the cepstral domain.[8–10] Linear or nonlinear spectral subtraction and statistical filtering approaches have also been used to extract noise-resistant features for recognition purposes.[11–15]

Wavelet transform has been proved to possess de-noising capability, particularly for Gaussian white noise.[16,17] This is because noise normally contributes to the wavelet coefficients evenly, while the original signal only affects those coefficients with relatively large magnitudes. As shown in Fig. 1, by thresholding the wavelet coefficients, i.e., shrinking large coefficients or zeroing small coefficients, noise can be greatly suppressed without causing significant distortion of the original signal. The effectiveness of this approach has been shown in a number of applications for signal enhancement.[18–20] However, Moon and Hwang showed that simply thresholding the wavelet coefficients is not good enough for robust recognition of noisy speech.[21] Instead, they proposed to apply iterative Wiener filtering on the wavelet coefficients prior to the inverse wavelet transform.[21] Although this leads to a significant improvement in the recognition accuracy for noisy speech, the iterative procedure presents a much higher computational requirement.

In this paper, we attempt to apply wavelet packet transform (WPT)-based Wiener filtering as a front-end feature extraction process for noisy speech recognition. A novel method is proposed to integrate the wavelet packet filter-bank architecture and the de-noising properties of the wavelet transform via Wiener filtering. WPT coefficients that characterize a set of preselected subbands are extracted, on a short-time basis, by the efficient multiresolution analysis (MRA) algorithm.[22] These coefficients form the basis of acoustic features for speech recognition using HMM. Wiener filter-based de-noising is then performed for those selected subbands at some early decomposition stages of the WPT. The filters are designed with special consideration of the spectral properties of each individual subband.

A brief introduction to wavelet transform and the MRA algorithm will be given in the next section. The general ap-

[a)]Author to whom correspondence should be addressed; electronic mail: tanlee@ee.cuhk.edu.hk

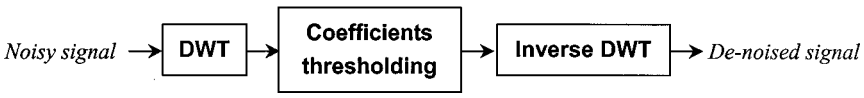


FIG. 1. Signal de-noising with DWT.

proach of wavelet de-noising by Wiener filtering will be described in Sec. III. The entire process for the extraction of de-noised MRA features is presented in Sec. IV. In Sec. V, the proposed features are evaluated and compared with other types of acoustic features in phone recognition experiments with the TIMIT database. In Sec. VI, a number of observations on the experimental results are discussed, in conjunction with an investigation on the acoustic distortion between the de-noised features and the clean ones. Finally, conclusions are given in Sec. VII.

II. WAVELET TRANSFORMS AND MULTIRESOLUTION ANALYSIS

Wavelet transform (WT) was first introduced as an analysis technique for nonstationary signals. In contrast to the conventional short-time Fourier transform (STFT) techniques, WT facilitates nonuniform or multiple resolutions in the time-frequency domain. This is particularly useful for speech analysis, which requires variable time and frequency localization for different sounds.

Given a discrete-time signal $s(k)$, the discrete wavelet transform (DWT) is defined as

$$\mathrm{DWT}(m,k)=2^{-m/2}\sum_{l} s(l)\psi(2^{-m}\cdot l-k), \tag{1}$$

where $\psi(\cdot)$ is called the wavelet basis function, m and k are integers, and $m\geqslant 0$. By varying m and k, the basis function is dyadically scaled and shifted. The computation of DWT essentially takes place every 2^m samples. A large value of m implies a large analysis window, resulting in a fine frequency representation and a coarse time localization.

DWT can be efficiently implemented by subband decomposition, as shown in Fig. 2. This algorithm is also commonly referred to as multiresolution analysis (MRA).[22] At each stage of decomposition, the signal is split into two half-band components. This is done with a pair of quadrature-mirror filters (QMF), followed by subsampling by a factor of 2. The QMFs, denoted by $h(k)$ and $g(k)$, are FIR high-pass and low-pass filters, respectively. Their subsampled outputs, $d_m(k)$ and $a_m(k)$, are referred to as the ''detail'' wavelet coefficients and the ''approximation'' wavelet coefficients, respectively. As m increases, the lower subbands are further decomposed, resulting in better frequency localization.

The MRA algorithm can be generalized in the way that decomposition is performed for both lower and higher frequency subbands (see Fig. 3). As a consequence, it becomes more flexible to configure the combination of subbands so as to satisfy the requirements of different applications. This operation is known as the wavelet packet transform (WPT). Figure 4 illustrates the subbands' distribution at the first two stages of WPT. In both figures, $d_{ij}(k)$'s denote the MRA outputs at the ith decomposition stage, where $j=0,1,...,2^i-1$.

III. WAVELET DE-NOISING BY WIENER FILTERING

A. Wiener filtering in the frequency domain

Wiener filtering has been widely applied to suppress additive noise, as illustrated in Fig. 5. Given a signal corrupted with additive white Gaussian noise (AWGN)

$$y(k)=s(k)+n(k), \tag{2}$$

where $y(k)$ denotes the corrupted signal, $s(k)$ the original signal, and $n(k)$ the AWGN. Then, the frequency response of a noncausal Wiener filter is given by[23]

$$H(\omega)=\frac{E[P_s(\omega)]}{E[P_s(\omega)]+E[P_n(\omega)]}, \tag{3}$$

where $E[P_s(\omega)]$ and $E[P_n(\omega)]$ are the estimated signal and noise power at a particular frequency ω, respectively.

Let $\hat{S}(\omega)$ and $Y(\omega)$ denote the de-noised and corrupted signal spectrum, respectively. Then, the de-noising operation can be expressed as

$$\hat{S}(\omega)=Y(\omega)H(\omega). \tag{4}$$

B. Wiener filtering in the wavelet domain

Wiener filtering can also be performed in the wavelet domain.[21] More precisely, for each subband of interest, a filter is designed to de-noise the WPT coefficients, based on the statistical properties of noise and speech. Given the linearity of WPT, the Wiener filter gain can be approximated by

$$H(i)=\frac{E[P_s(i)]}{E[P_s(i)]+E[P_n(i)]}, \tag{5}$$

or, equivalently, by

$$H(i)=\frac{\Re_o(i)}{\Re_o(i)+1}, \tag{6}$$

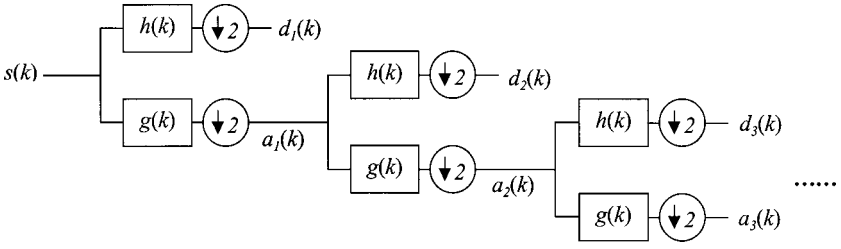


FIG. 2. The multiresolution analysis (MRA) algorithm for subband decomposition.

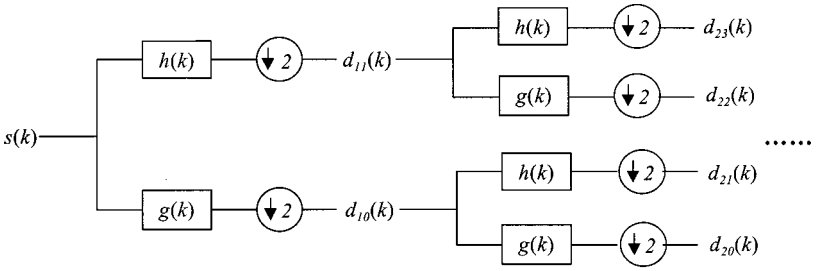


FIG. 3. Wavelet packet transform by subband decomposition.

where $E[P_s(i)]$ and $E[P_n(i)]$ represent the estimated signal and noise power in the ith subband, respectively, and $\Re_o(i)=E[P_s(i)]/E[P_n(i)]$ is the signal-to-noise ratio (SNR). The wavelet de-noising is carried out by

$$\hat{d}_i = H(i)\cdot c_i, \tag{7}$$

where c_i and $\hat{d}_i$ are the corrupted and de-noised WPT coefficients, respectively. $H(i)$ can be considered as the "average" Wiener filter gain over a particular subband, in contrast to $H(\omega)$ in Eq. (3), which is defined for every value of frequency ω.

It is noted that generalized Wiener filtering, when being applied to the spectral domain, is very similar to spectral subtraction. Both methods attempt to remove noise in the same way. The proposed de-noising technique differs from the conventional spectral subtraction approach in that Wiener filtering is applied to the WPT domain on a subband basis.

C. Estimation of speech and noise power

In our application, the Wiener gain has to be computed for each short-time frame from which a de-noised feature vector is derived. Given a speech utterance, the noise power $E[P_n(i)]$ is estimated from the first *five* frames. It is assumed that the additive noise is stationary over the entire utterance. For a particular frame, the speech power $E[P_s(i)]$ can be obtained by

$$E[P_s(p,i)] = P_y(p,i) - E[P_n(i)], \tag{8}$$

where p indexes the time frames and $P_y(p,i)$ denotes the subband power of the corrupted signal. For real speech signals, $P_y(p,i)$ exhibits rapid fluctuation and so does the resultant Wiener gain. Therefore, the following smoothing process is applied:

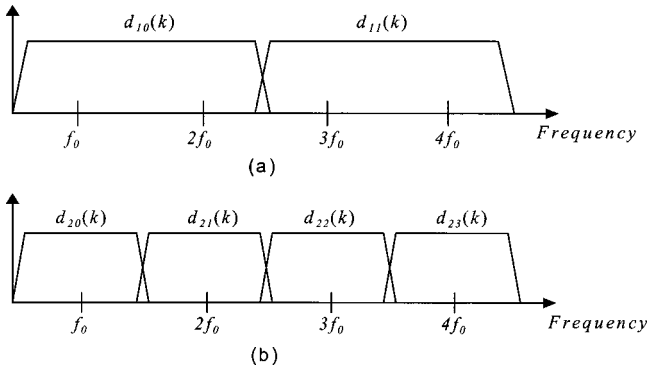


FIG. 4. Subbands characterized by the outputs of WPT at (a) stage 1 and (b) stage 2.

$$\Re_s(p,i) = (1-\alpha)\Re_o(p,i) + \alpha\frac{H(p-1,i)^2 P_y(p-1,i)}{E[P_n(i)]}, \tag{9}$$

where $\Re_s(p,i)$ and $\Re_0(p,i)$ are the postsmoothing and the original SNR, respectively. The parameter α varies between 0 and 1 to control the degree of smoothness. It is empirically determined to be 0.85 in all of our recognition experiments.

The smoothed subband Wiener gain for the pth frame can then be written as

$$H(p,i) = \frac{\Re_s(p,i)}{\Re_s(p,i)+1}. \tag{10}$$

This method is based on the noise suppression technique proposed by Ephraim and Malah[24] and generalized by Cappe.[25] The effect of smoothing is illustrated by an example utterance, as depicted in Fig. 6.

IV. EXTRACTION OF ACOUSTIC FEATURES WITH EMBEDDED DE-NOISING

A. The mel-scaled wavelet packet filter-bank

As described in Sec. II, the MRA algorithm splits the signal into half-bands in successive stages. This allows much flexibility in selecting frequency bands to meet the requirements of specific applications. For speech recognition, spectral analysis is usually performed with the mel-scale filter-bank with a view to approximating human auditory response.[26] In this work, we propose to use the wavelet packet tree, as shown in Fig. 7. It approximates a mel-scale filter-bank with 22 filters. The signal bandwidth is assumed to be 8 KHz.

B. Wiener filtering at selected subbands

Instead of performing filtering individually for all of the 22 subbands, we select only four of them to apply the Wiener de-noising operation to, as illustrated in Fig. 7. There are a number of reasons for such a selection. First, Wiener filtering is computationally intensive. To make the proposed method practical and useful, we should avoid an excessive number of Wiener operations. Second, the number of coefficients for

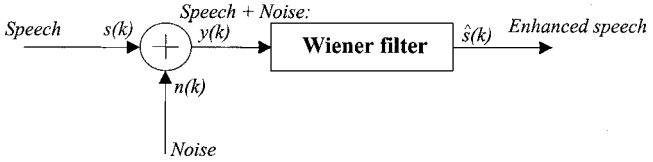


FIG. 5. Wiener filtering for speech enhancement.

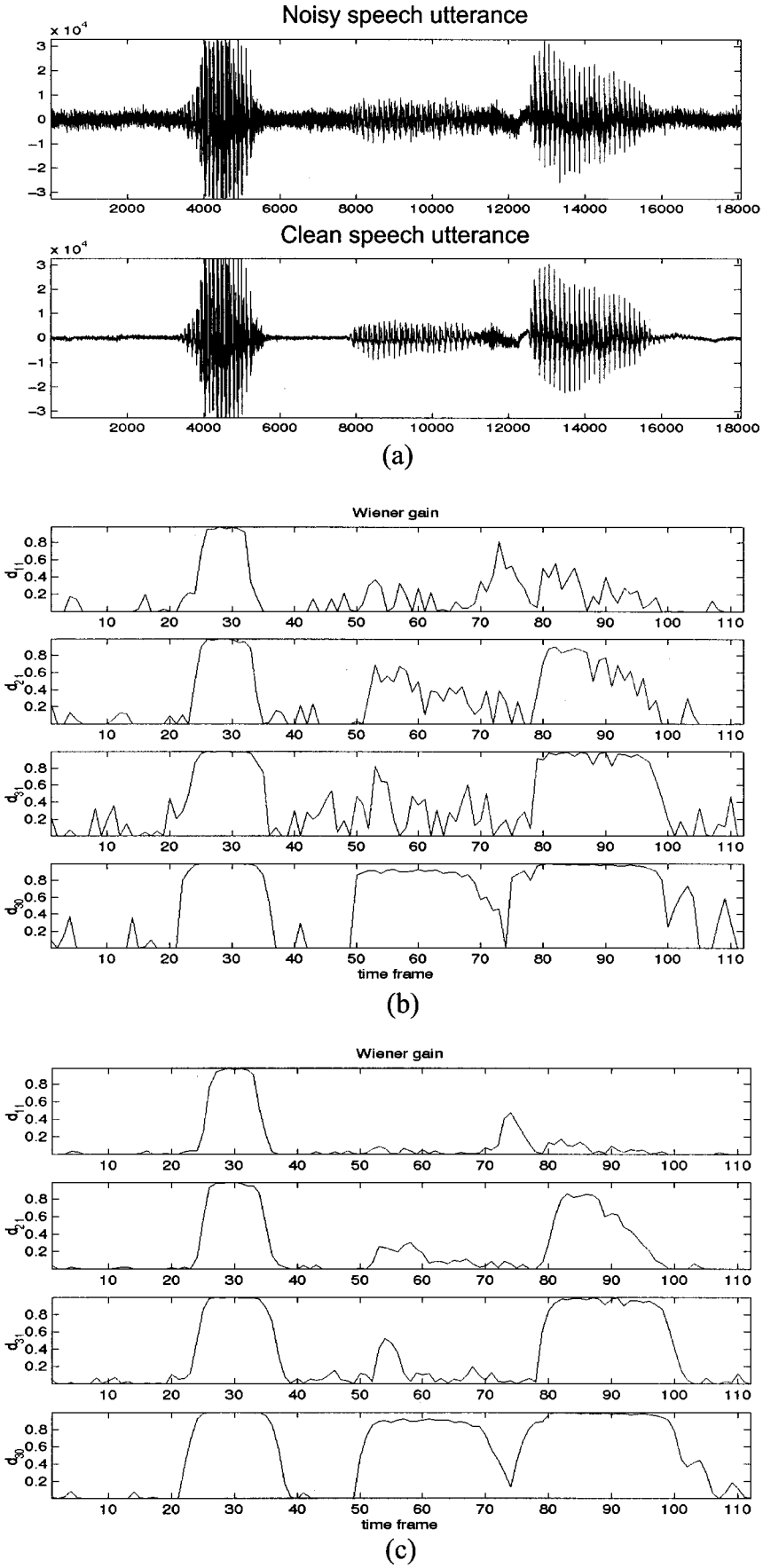


FIG. 6. The effect of smoothing the Wiener gain: (a) noisy and clean speech utterances; (b) the original subband Wiener gains; (c) the smoothed Wiener gains.

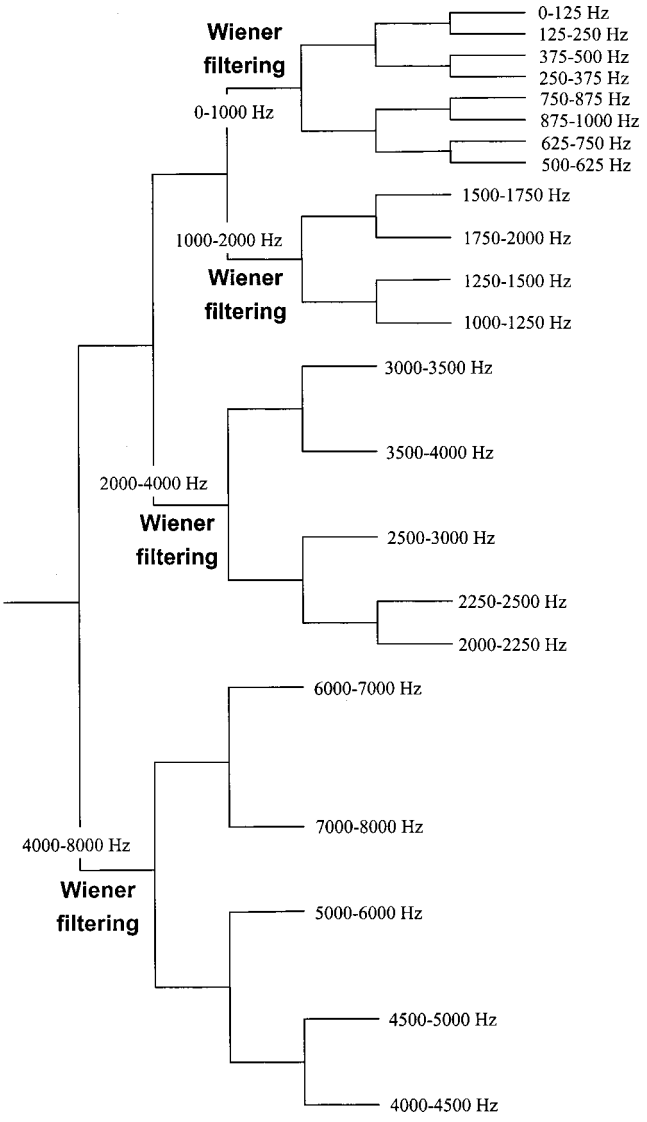


FIG. 7. The mel-scaled wavelet packet tree with embedded subband denoising by Wiener filtering.

representing a subband is halved by each decomposition operation in the MRA. That is, fewer coefficients are used to represent finer subbands. It is not desirable to perform Wiener filtering at later stages of decomposition, since a small set of coefficients generally does not provide a reliable estimation of signal and noise power, as described in Sec. III. Third, spectral analysis of speech signals requires higher resolution in lower frequency ranges. This explains why Wiener filtering is applied to the bands 0–1000 Hz and 1000–2000 Hz individually, but the band 4000–8000 Hz is processed as a whole.

Consonant sounds usually span the high-frequency subbands d_{11} (4 K–8 KHz) and d_{21} (2 K–4 KHz). They generally have fairly low signal power in comparison with vowel sounds. It is possible that the estimated subband noise power overwhelms the signal power. Therefore, we use a fraction of the estimated subband noise power to calculate the postsmoothing SNR in d_{11} and d_{21}, i.e.,

$$\Re_o(p,i)=\frac{P_y(p,i)-\beta E[P_n(i)]}{\beta E[P_n(i)]}=\frac{P_y(p,i)}{\beta E[P_n(i)]}-1, \quad (11)$$

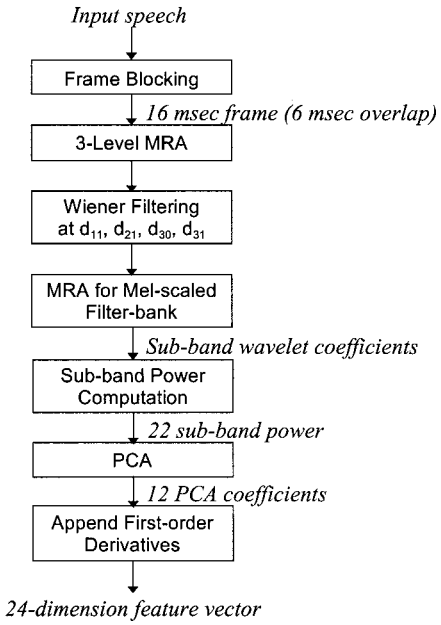


FIG. 8. The overall process of extracting the MRA de-noised acoustic feature vector.

$$\Re_s(p,i)=(1-\alpha)\Re_o(p,i)+\alpha\frac{H(p-1,i)^2P_y(p-1,i)}{\beta E[P_n(i)]}. \quad (12)$$

In our recognition experiments, the parameter β was empirically set at 0.5 and 0.7 for d_{11} and d_{21}, respectively.

C. Derivation of the feature vector

For each subband in Fig. 7, the subband power is computed by

$$E_{ij}=\log\left[\frac{1}{N(i,j)}\sum_{k=0}^{N(i,j)-1}d_{ij}^2(k)\right], \quad (13)$$

i.e., the power computed at node j of decomposition level i. Here, $d_{ij}(k)$ denotes the output coefficients, $N(i,j)$ is the total number of coefficients at that node, and E_{ij}'s essentially characterize the distribution of signal power in the mel-scaled frequency bands.

In this study, continuous-density HMMs with diagonal covariance matrices are employed to model the acoustic input. This requires the individual feature parameters to be statistically uncorrelated. Principal component analysis (PCA) has been applied to the subband power parameters for dimension reduction and decorrelation. PCA is an orthonormal transformation that provides a linear mapping from high dimension to lower dimension feature space, subject to the criterion of minimum mean-square error (MMSE).[27,28] The effectiveness of PCA in pattern recognition is due to its ability to decorrelate feature parameters and relegate most of the random structures to trailing components while extracting systematic patterns to leading ones.

The overall process of feature extraction is shown in Fig. 8. The input utterance is divided into overlapping frames. The frame size is 16 ms and the frame shift is 10 ms.

TABLE I. HMM definition for the phone recognition experiments.

Total no. of HMMs	61 (context-independent phone models)
No. of states per HMM	3
HMM topology	Left-to-right without state skipping
No. of Gaussian mixtures per state	8
Covariance matrix	Diagonal

For each frame, a 24-dimensional feature vector is formed, based on the PCA outputs. It contains 12 static PCA coefficients and their first-order derivatives.

V. PHONE RECOGNITION EXPERIMENTS

The proposed method was evaluated in phone recognition experiments with the TIMIT speech database.[29] Phone recognition is of fundamental importance in the entire ASR process. It reveals the effectiveness of acoustic features, which is indeed the focus of this study. Furthermore, TIMIT is a generic database with balanced and comprehensive acoustic–phonetic coverage, and great speaker variability. It provides a suitable platform for an all-sided investigation of phone recognition.

In addition to the de-noised MRA features, a number of commonly used feature types were tested on the same task to facilitate performance comparison. These feature types are divided into two categories, as follows:

Category A — Features without noise subtraction

(i) *MFCC:* The standard mel-frequency cepstral coefficients. The filter-bank analysis uses a Hamming window and 33 triangular filters.

(ii) *FBANK_PCA:* Filter-bank analysis followed by PCA. The procedure is very similar to that of the *MFCC*, except that the filter-bank outputs are postprocessed by the PCA instead of the DCT.

(iii) *WPT_PCA:* The MRA features without subband Wiener filtering. Daubechies 12-tap FIR filter with ''least asymmetric'' compactly supported wavelets is adopted.[30]

Category B — Features with noise subtraction

(i) *MFCC_CMN:* Same as the *MFCC* in (1), except that utterance-based cepstral mean normalization (CMN) is applied.[31]

TABLE II. Phone recognition rates for different feature types. Those in bold fonts indicate the best performance under the respective noisy conditions.

	Feature types	Test data		
		Clean	20 dB	10 dB
Category A: w/o noise subtraction	**MFCC**	60.25%	37.85%	20.02%
	FBANK PCA	**63.39%**	37.26%	24.80%
	WPT_PCA	58.05%	36.76%	24.86%
Category B: w/ noise subtraction	**MFCC_CMN**	60.09%	43.25%	28.74%
	WPT_SMSNR	57.00%	43.85%	29.61%
	WPT_SMSNR_PARTN	57.17%	**46.17%**	**32.08%**

TABLE III. Phone-class-based recognition rates under the condition of 10-dB SNR.

Phone class	Occurrence count in class	Feature types		
		MFCC	**MFCC_CMN**	**WPT_SMSNR_PARTN**
Vowel	16 578	21.57%	47.20%	38.27%
Semivowel & glide	5 888	12.34%	20.39%	35.20%
Nasal	4 436	0.34%	10.92%	30.49%
Stop	7 266	14.68%	14.38%	14.56%
Affricate	554	15.34%	7.94%	51.81%
Fricative	6721	42.09%	19.49%	32.44%

(ii) *WPT_SMSNR:* The de-noised MRA features obtained, as shown in Fig. 8. The subband Wiener gains are smoothed by Eqs. (9) and (10), with the parameter $\alpha = 0.85$.

(iii) *WPT_SMSNR_PARTN:* Same as *WPT_SMSNR*, except that the smoothing strategy for the high-frequency bands is changed, as in Eqs. (11) and (12) where $\beta = 0.5$ for d_{11} and $\beta = 0.7$ for d_{21}.

For all of these feature types, a frame size of 16 ms is used. The feature vectors have the same dimension of 24, containing 12 static features together with their first-order derivatives. The training data contain 3696 utterances from 462 speakers while the test data contain 1344 utterances from 168 speakers. In total there are 41 443 occurrences of phones in the test data. A total of 61 phonetic labels was used in TIMIT. These include all English phones in the classes of stops, fricatives, semivowels, glides, and vowels. For each phone, a context-independent mode was built with a continuous-density HMM, as shown in Table I.

Noise-free training data were utilized to train the HMMs. Subsequently, the resultant ''clean'' models were used to recognize clean and noisy speech with 20-dB and 10-dB SNR, respectively. The noisy speech data were generated by adding stationary white Gaussian noise. The recognition performance was measured by performing an optimal string match between the recognized phone sequence and the correct one. The phone recognition rate, as listed in Table II, is defined as the percentage of matched phones over the total number of phone occurrences. As a matter of fact, insertion errors are inevitable in phone recognition since the occurrence of a phone is completely determined by acoustic evidence, without any linguistic constraints. However, in the Viterbi search for the optimal phone sequence, it is possible to trade the percentage of insertion errors with the percentage of matched phones by adjusting a penalty factor.[31] In our experiments, this penalty factor was selected so that the insertion errors were kept at a fixed percentage of 10% (or approximately 4150 in number).

In many previous recognition experiments with TIMIT, a number of confusing phone groups were identified in the sense that within-group confusions were not counted as recognition errors.[32–34] Following this convention, the experi-

mental results in Table II were based on only 39 distinguishable phones.

Table III compares the phone-class-based recognition rates given by *MFCC*, *MFCC_CMN*, and *WPT_SMSNR_PARTN* in the case of 10-dB SNR. The class-based recognition rate is calculated as the percentage of correctly recognized phones in a particular phone class over the total number of phone occurrences in that class. It indicates the effectiveness of the acoustic features for different types of speech sounds.[32]

VI. DISCUSSION

From the experimental results in Tables II and III, the following can be observed:

(1) For clean speech, the feature types of category A generally performed slightly better than those of category B, with the exception of the *MFCC_CMN*. This reveals that the de-noising techniques in general may trade off some acoustic properties of the speech signals and hence degrade the recognition performance.
(2) In the presence of stationary AWGN, the performance of all feature types of category A dropped drastically from about 58%–63% to 36%–37% and 20%–24% in the cases of 20-dB and 10-dB SNR, respectively. In contrast, the category B features did not suffer that much performance degradation since they had incorporated different techniques for noise removal.
(3) For noisy speech, the proposed MRA features consistently outperformed the other feature types. The use of the *WPT_SMSNR_PARTN* resulted in phone recognition rates of 32.08% and 46.17% at the SNRs of 10 dB and 20 dB, respectively, in comparison to 29.21% and 43.25% attained by the commonly used robust features, the *MFCC_CMN*.
(4) As shown in Table III, the *WPT_SMSNR_PARTN* was extremely robust for recognizing semivowels, nasals, and affricates in noisy speech. It obviously outperformed the *MFCC_CMN* for almost all phone classes except vowels.
(5) *WPT_SMSNR_PARTN* consistently performed better than *WPT_SMSNR* under different SNR conditions. This demonstrates the effectiveness of the proposed special treatments for high-frequency bands in the design of Wiener filters.

Furthermore, we attempted to investigate the effectiveness of the proposed de-noising method by measuring the distortion between the de-noised feature parameters and those extracted from clean speech. Let $\{z_1, z_2,\ldots,z_{12},\Delta_1,\Delta_2,\ldots,\Delta_{12}\}$ and $\{\tilde{z}_1,\tilde{z}_2,\ldots,\tilde{z}_{12},\tilde{\Delta}_1,\tilde{\Delta}_2,\ldots,\tilde{\Delta}_{12}\}$ denote the ''clean'' and the de-noised feature vectors, respectively. z_n's and Δ_n's represent static and dynamic features, respectively. The Euclidean distances are then computed for static and dynamic features separately, i.e.,

$$d_{\text{static}}=\sqrt{\sum_{n=1}^{12}(z_n-\tilde{z}_n)^2}, \tag{14}$$

TABLE IV. Distortion measures for different feature types under the condition of 10-dB SNR.

	d_{static}	d_{dynamic}
FBANK_PCA	12.58	1.69
WPT_PCA	16.88	2.55
WPT_SMSNR	10.55	2.21
WPT_SMSNR_PARTN	8.89	2.19

$$d_{\text{dynamic}}=\sqrt{\sum_{n=1}^{12}(\Delta_n-\tilde{\Delta}_n)^2}. \tag{15}$$

Eight utterances from a particular speaker (speaker code: FAKS0; dialect region: DR1) in TIMIT were used to evaluate the feature distortion. The utterances covered all phonetic classes. Table IV shows the mean values of d_{static} and d_{dynamic} for the four feature types that involve PCA. As for the *MFCC* and the *MFCC_CMN*, the distance defined by Eqs. (14) and (15) is given a totally different interpretation from that given to the PCA derived features.

As shown in Table IV, the de-noised features, namely *WPT_SMSNR* and *WPT_SMSNR_PARTN*, have much smaller distortion than those without de-noising. In other words, the proposed technique is helpful to alleviate the acoustic mismatch between clean training data and noisy test data. The de-noising process seems to be more effective for static features than for dynamic features. This might be due to the fact that the dynamic features are relatively smaller in value and hence more susceptible to additive noise.

VII. CONCLUSION

A new method of extracting robust acoustic features has been proposed and evaluated. Wiener filtering was embedded into the wavelet packet filter-bank framework to generate de-noised subband features for the recognition of speech with additive Gaussian white noise. Wiener filtering was applied only to a small number of selected subbands so as to keep the additional computation cost minimal. Moreover, different filter design strategies were employed for higher frequency bands to make the consonant features more prominent. The de-noised MRA features exhibited good robustness to stationary AWGN in phone recognition experiments. It was also shown that the de-noising process effectively reduces the distortion between the noise-corrupted features and clean speech features. Research in this direction is still in its early stages and much further work is required before the proposed method will be useful in practical speech recognition systems. In particular, noise types other than stationary AWGN need to be considered.

ACKNOWLEDGMENTS

This work is partly supported by a research grant from the Hong Kong Research Grants Council. The authors would like to thank Y. W. Wong for his technical assistance.

[1] J.-C. Junqua and J.-P. Haton, *Robustness in Automatic Speech Recognition: Fundamentals and Applications* (Kluwer Academic, Dordrecht, 1996).
[2] A. Nadas, D. Nahamoo, and M. A. Picheny, ''Speech recognition using

noise-adaptive prototypes,'' IEEE Trans. Acoust., Speech, Signal Process. **37**, No. 10, 1495–1503 (1989).
[3] J.-T. Chien, L.-M. Lee, and H.-C. Wang, ''Noisy speech recognition by using variance adapted hidden Markov models,'' IEE Electron. Lett. **31**, No. 18, 1555–1556 (1995).
[4] A Varga and R. Moore, ''Hidden Markov model decomposition of speech and noise,'' in Proceedings of 1990 International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, pp. 845–848, April 1990.
[5] M. J. F. Gales and S. J. Young, ''HMM recognition in noise using parallel model combinations,'' in Proceedings of 1993 European Conference on Speech Communications and Technology, pp. 837–840, September 1993.
[6] J.-C. Junqua, H. Wakita, and H. Hermansky, ''Evaluation and optimization of perceptually-based ASR front-end,'' IEEE Trans. Speech Audio Process. **1**, No. 1, 39–48 (1993).
[7] D. Mansour and B. H. Juang, ''The short-time modified coherence representation and noisy speech recognition,'' IEEE Trans. Acoust., Speech, Signal Process. **37**, No. 6, 795–804 (1989).
[8] A. Acero and R. M. Stern, ''Environmental robustness in automatic speech recognition,'' in Proceedings of 1990 International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, pp. 849–852, April 1990.
[9] F.-H. Liu, R. M. Stern, A. Acero, and P. J. Moreno, ''Environmental normalization for robust speech recognition using direct cepstral comparison,'' in Proceedings of 1994 International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, pp. 61–64, April 1994.
[10] M. Morishima, T. Isobe, and J. Takahashi, ''Phonetically adaptive cepstrum mean normalization for acoustic mismatch compensation,'' in Proceedings of 1997 IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 436–441, December 1997.
[11] S. F. Boll, ''Suppression of acoustic noise in speech using spectral subtraction,'' IEEE Trans. Acoust., Speech, Signal Process. **27**, No. 2, pp. 113–120 (1979).
[12] P. Lockwood, J. Boudy, and M. Blanchet, ''Nonlinear spectral subtraction (NSS) and hidden Markov models for robust speech recognition in car noise environments,'' in Proceedings of 1992 International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 265–268, March 1992.
[13] A. D. Berstein and I. D. Shallom, ''A hypothesized Wiener filtering approach to noisy speech recognition,'' in Proceedings of 1991 International Conference on Acoustics, Speech, and Signal Processing, pp. 913–916, April 1991.
[14] S. V. Vaseghi and B. P. Milner, ''Noise adaptive hidden Markov models based on Wiener filters,'' in Proceedings of 1993 European Conference on Speech Communications and Technology, pp. 1023–1026, September 1993.
[15] B. H. Juang, ''Speech recognition in adverse environments,'' in Comput. Speech Lang. **5**, 275–294 (1991).
[16] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer* (Prentice-Hall, New York, 1998).
[17] D. L. Donoho, ''De-noising by soft-thresholding,'' IEEE Trans. Inf. Theory **41**, No. 3, 613–627 (1995).
[18] N. A. Whitmal, J. C. Rutledge, and J. Cohen, ''Wavelet-based noise reduction,'' in Proceedings of 1995 International Conference on Acoustics, Speech, and Signal Processing, Vol. 5, pp. 3003–3006, May 1995.
[19] H. A. Kestler *et al.*, ''De-noising of high-resolution ECG signals by combining the discrete wavelet transform with the Wiener filter,'' in Proceedings of Computers in Cardiology 1998, pp. 233–236 (1998).
[20] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, ''Signal de-noising using adaptive Bayesian wavelet shrinkage,'' in Proceedings of 1996 IEEE-SP international Symposium on Time-Frequency and Time-Scale Analysis, pp. 225–228, June 1996.
[21] S. Moon and J. -N. Hwang, ''Noisy speech recognition via wavelet transform enhancement,'' in Proceedings of the Twenty-sixth Annual Asilomar Conference on Signals, Systems and Computers, Vol. 2, pp. 1086–1090, October 1992.
[22] S. G. Mallat, ''A theory for multi-resolution signal decomposition: The wavelet representation,'' IEEE Trans. Pattern Anal. Mach. Intell. **11**, No. 7, 674–693 (1989).
[23] J. S. Lim and A. V. Oppenheim, ''All-pole modeling of degraded speech,'' IEEE Trans. Acoust., Speech, Signal Process. **26**, No. 3, 197–210 (1978).
[24] Y. Ephrain and D. Malah, ''Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,'' IEEE Trans. Acoust., Speech, Signal Process. **32**, No. 6, 1109–1121 (1984).
[25] O. Cappe, ''Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor,'' IEEE Trans. Speech Audio Process. **2**, No. 2, 345–349 (1994).
[26] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition* (Prentice-Hall, New York, 1993).
[27] K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Academic, New York, 1972)
[28] P. J. Huber, ''Projection pursuit,'' Ann. Stat. **13**, No. 2, 435–475 (1985).
[29] The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus (CD-ROM documentation), 1990.
[30] I. Daubechies, *Ten Lectures on Wavelets* (SIAM, Philadelphia, 1992).
[31] S. Young, D. Ollason, V. Valtchev, and P. Woodland, *The* HTK *Book (for* HTK *Version 2.1*), Entropic Research Laboratory (1997).
[32] K.-F. Lee and H.-W. Hon, ''Speaker-independent phone recognition using hidden Markov models,'' IEEE Trans. Acoust., Speech, Signal Process. **37**, No. 11, 1641–1648 (1989).
[33] V. V. Digalakis, M. Ostendorf, and J. R. Rohlicek, ''Fast algorithms for phone classification and recognition using segment-based models,'' IEEE Trans. Signal Process. **40**, No. 12, 2885–2896 (1992).
[34] A. J. Robinson, ''An application of recurrent nets to phone probability estimation,'' IEEE Trans. Neural Netw. **5**, No. 2, 298–305 (1994).

Cross-spectral methods for processing speech

Douglas J. Nelson[a)]
Department of Defense, Ft. Meade, Maryland 20755

(Received 22 March 2000; revised 11 October 2000; accepted 19 July 2001)

We present time-frequency methods which are well suited to the analysis of nonstationary multicomponent FM signals, such as speech. These methods are based on group delay, instantaneous frequency, and higher-order phase derivative surfaces computed from the short time Fourier transform (STFT). Unlike more conventional approaches, these methods do not assume a locally stationary approximation of the signal model. We describe the computation of the phase derivatives, the physical interpretation of these derivatives, and a re-mapping algorithm based on these phase derivatives. We show analytically, and by example, the convergence of the re-mapping to the FM representation of the signal. The methods are applied to speech to estimate signal parameters, such as the group delay of a transmission channel and speech formant frequencies. Our goal is to develop a unified method which can accurately estimate speech components in both time and frequency and to apply these methods to the estimation of instantaneous formant frequencies, effective excitation time, vocal tract group delay, and channel group delay. The proposed method has several interesting properties, the most important of which is the ability to simultaneously resolve all FM components of a multicomponent signal, as long as the STFT of the composite signal satisfies a simple separability condition. The method can provide super-resolution in both time and frequency in the sense that it can simultaneously provide time and frequency estimates of FM components, which have much better accuracy than the Heisenberg uncertainty of the STFT. Super-resolution provides the capability to accurately ''re-map'' each component of the STFT surface to the time and frequency of the FM signal component it represents. To attain high resolution and accuracy, the signal must be jointly estimated simultaneously in time and frequency. This is accomplished by estimating two surfaces, which are essentially the derivatives of the STFT phase with respect to time and frequency. To avoid phase ambiguities, the differentiation is performed as a cross-spectral product. [DOI: 10.1121/1.1402616]

PACS numbers: 43.72.Ar, 43.72.Lc [DOS]

I. INTRODUCTION

Speech and other biological signals pose an interesting signal processing dilemma. The signal may be represented as the sum of resonances which result when the vocal tract is excited by a combination of frication and a sequence of glottal pulses, but nothing in the signal is stationary. The repetition frequency, $F0$, of the glottal pulses can vary by an octave or more for any speaker, and there may be considerable instability on a pulse-by-pulse basis. There is considerable FM modulation of the resonant frequencies or formants induced by movement of the tongue, lips, and other components of the vocal tract. In addition, there is considerable loss in the vocal tract, resulting in a rapid AM modulation of the formants.[1]

There have been several excursions into spectral phase processing of signals. Most of these processes are based on the instantaneous frequency and group delay functions. These functions are not new concepts, having appeared in the literature more than 30 years ago. The group delay was discussed by Gibbs in his 1970 paper,[2] and the instantaneous frequency goes back at least to the Thomas book published in 1969.[3] The harmonic-rejecting correlation function (HRF)[4,5] and crosspower spectrum[6,5] were developed in the 80's to accurately estimate the frequencies of complex pulse trains and stationary tones. Applications to speech include the instantaneous-frequency-distribution (IFD), also developed in the 80's and applied to an early recognition effort.[7] In the 90's, instantaneous frequency has been used to estimate carrier offset of mistuned single sideband AM from the observed pitch harmonics of the transmitted speech signal[8] and to estimate pitch frequencies.[9,10] Our goal is to develop a unified method which can accurately estimate speech components in both time and frequency and to apply these methods to the estimation of instantaneous formant frequencies, effective excitation time, vocal tract group delay, and channel group delay. We develop an analysis method based on the gradient of the short time Fourier transform (STFT) phase. The proposed method has several interesting properties, the most important of which is the ability to simultaneously resolve all FM components of a multicomponent signal, as long as the STFT of the composite signal satisfies a simple separability condition. The method can provide super-resolution in both time and frequency in the sense that it can simultaneously provide time and frequency estimates of FM components, which have much better accuracy than the Heisenberg uncertainty of the STFT. Super-resolution provides the capability to accurately ''re-map'' each component of the STFT surface to the time and frequency of the FM signal component it represents. To attain high resolution and accuracy, the signal must be jointly estimated simultaneously

a)Electronic mail: waveland@erols.com, www.wavelandplantation.com

in time and frequency. This is accomplished by estimating two surfaces, which are essentially the derivatives of the STFT phase with respect to time and frequency. To avoid phase ambiguities, the differentiation is performed as a cross-spectral product.

In addition to the signal time and frequency estimation, a second-derivative test is developed which indicates the local convergence of the re-mapped surface. By modeling the speech signal as a short period of excitation followed by a nearly steady state damped resonance, it is shown that the second derivative test for convergence of the re-mapped surface provides a convenient test for determining which specific points on the STFT surface represent a response to excitation or resonance. This second-derivative test is used to isolate resonance related STFT components, from which it is demonstrated that accurate instantaneous formant frequencies may be estimated. In addition, the excitation components are isolated, and the system group delay function is estimated from the re-mapped excitation surface. By analyzing speech which has not been subjected to a transmission channel, the group delay of the vocal tract is estimated, and it is demonstrated that the group delay of the vocal tract is essentially constant in the sense that the entire spectrum of the voiced speech signal apparently responds to excitation at the same time. With this condition, it is demonstrated that the channel group delay function may be estimated from the re-mapped STFT surface.

This paper consists of three main sections. In Sec. II, we develop the cross-spectral phase-based re-mapping of the STFT surface. In Sec. II, we define the various time-frequency (TF) surfaces, the STFT re-mapping, and the separability condition under which nonlinear operations, such as cross-spectral products, behave linearly. In Sec. II A, we provide interpretations of the local group delay, the channelized instantaneous frequency, the mixed partial derivative surface, and the STFT re-mapping. In Secs. II B and II C, we examine the re-mapping properties of the STFT of a linear-FM (chirp) signal. We provide an analytic proof of the convergence of the re-mapping of the STFT surface to the functional-FM representation of the chirp. It is also proven that re-mapping of the STFT surface converges for sine waves and impulses in time, which are the limiting cases of linear-FM signals. In addition, the convergence under the re-mapping extends to signals which may be approximated as piecewise linear-FM signals. In Sec. III, the experimental simulations are presented which verify the theoretical results of Sec. II. In Sec. IV, we apply the methods developed in Sec. II to the analysis of speech. We start with a short discussion of the general channel model and the single pulse excitation model of speech. In Secs. IV B and IV C, we develop models for the received speech signal at excitation and resonance. Under these models, it is shown analytically that excitation and resonance have orthogonal STFT phase properties. Excitation affects only the local STFT group delay properties, and steady state resonance affects only the STFT channelized instantaneous frequency properties. Section IV D is an analytic development of the mixed partial phase derivatives as a resonance/excitation test. In Sec. IV E, the excitation and resonance models are validated by applying the cross-spectral methods to samples of speech signals. It is verified that the re-mapped STFT representation of speech is consistent with the resonance/excitation speech model and that the cross-spectral methods may be used effectively on speech. Finally, in Sec. IV F, we apply the methods to blindly estimate the channel group delay function from the re-mapped STFT of the signal at excitation. The estimated channel group delay is validated by comparison of the re-mapped and equalized time frequency surface with the re-mapped surface of the original signal which was not subjected to the channel.

Spectral phase is used extensively in this paper to recover speech signal information. In Sec. V, we present a brief case, based on the literature, that spectral phase is also important in human hearing, although it is largely ignored by the speech research community. The intent of this section is not to spark controversy but to demonstrate that it is possible that the spectral phase is preserved in the auditory process and that it is possible to use spectral phase-based methods to approximate human performance. This is demonstrated by a simple example for the case of frequency discrimination.

II. STFT, PHASE DERIVATIVES AND CROSS-SPECTRAL METHODS

In this section, we present a method based on the phase gradient of the STFT, which provides high resolution in both time and frequency. The method is based on the use of the derivatives, with respect to both time and frequency, of the STFT phase and provides a re-mapping of the STFT surface. The phase derivatives are computed as the arguments of cross-spectral surfaces, which are the product of the STFT of the signal and the complex conjugate of the STFT of the signal delayed in time and/or frequency.

While phase derivatives may be estimated by other methods, the primary advantage of the cross-spectral representation is that there is no need to phase unwrap the STFT surface to resolve the discontinuities of the principle-value representation of the argument function. In addition, each point of the cross-spectral surfaces has the STFT energy at that point encoded as magnitude, and the derivative of the STFT phase encoded as phase.

In the cross-spectral processes, the spectral magnitude is not used directly in the computation of frequency or group delay estimates, but it does play an important role in calculating average or mean frequencies. Spectral amplitude is also used as a criterion for isolation and estimation of the individual FM signal components of multicomponent signals. This criterion is stated as a separability condition under which these individual signal components do not interfere with each other on the STFT surface.

We describe the computation of the phase derivatives, the physical interpretation of these derivatives, and the re-mapping algorithm. We also show analytically, and by example, that the STFT surface of a linear-FM signal (chirp) is re-mapped to the line in the TF plane which is the functional-FM representation of that chirp. That is, for each point (ω_0, T_0) on the STFT surface of an arbitrary linear-FM signal, there is some value of t_0 such that (ω_0, T_0) is re-mapped as

$$(\omega_0, T_0) \rightarrow \left(\frac{d}{dt} \arg\{f(t_0)\}, t_0 \right). \tag{1}$$

Since the re-mapping is a local property, the STFT surface of an arbitrary FM modulated signal, which may be approximated by piecewise linear-FM signals, must be re-mapped to the curve in the TF plane which is the functional-FM representation of that FM signal. This means that, for FM signals, the re-mapping of the STFT effectively removes the blurring in both time and frequency and, therefore, provides super-resolution in both dimensions. In time, the blurring of the original STFT surface results from the fact that the Fourier transform must be computed as the integral of the signal over some time interval, making it difficult to estimate the signal timing to more accuracy than the length of that interval. In frequency, the blurring results from the fact that the frequency resolution of the Fourier transform is proportional to the reciprocal of the integration time of the Fourier transform. A standard estimate of the joint TF resolution is the Heisenberg uncertainty principle, which states, for the Fourier transform[11]

$$\Delta\omega \Delta T \geqslant \tfrac{1}{2}, \tag{2}$$

where $\Delta\omega$ and ΔT are the bandwidth and time duration of the observation. Re-mapping provides more accuracy and precision than the Heisenberg limit, but it requires two factors, which are not assumed in the Heisenberg limit. These are separability and an AM/FM model of the signal. Separability will be seen to guarantee that the signal components are resolved, and the model assumption adds information, which is not assumed by the Heisenberg principle.

In the STFT, the Fourier transforms of the product of the signal $f(t)$ and a sequence of time translations of a (short) analysis window $h(-t)$ are computed. The STFT may therefore be represented as

$$F_h(\omega, T) = \int_{-\infty}^{\infty} f(t+T) h(-t) e^{-i\omega t}\, dt, \tag{3}$$

where we have represented the window as a function of $-t$ to represent the STFT as a convolution, and we have followed the matrix notation in representing the first coordinate by ω. The cross spectrum of two signals, $f(t)$ and $g(t)$, is

$$R_{fg}(\omega, T) = F_{h_1}(\omega, T) G^*_{h_2}(\omega, T), \tag{4}$$

and is easily seen to be the Fourier transform of the cross correlation of the signals $f(t)$ and $g(t)$.

Now, let us assume that we have a signal which may be written as the sum of AM and FM modulated components

$$f(t) = \sum_n f_n(t), \tag{5}$$

where $f_n(t) = A_n(t) e^{i(\Omega_n(t) + \phi_n)}$, $\omega_n(t) = (d/dt)\Omega_n(t)$ and both $A_n(t)$ and $\omega_n(t)$ are continuous and do not change so rapidly that the following separability condition is invalid. We will define the STFT surface to be separable at a point (ω_0, T_0) if there is at most one component f_n which has significant energy in a neighborhood of that point. That is, if $|F_n(\omega_0, T_0)|^2$ is large, then

$$|F_n(\omega, T)|^2 \gg \sum_{m \neq n} |F_m(\omega, T)|^2, \tag{6}$$

where $|\omega - \omega_0| < \epsilon$ and $|T - T_0| < \delta$ for some ϵ and δ, and we have dropped the "h" from the STFT notation for simplicity. Separability is an important concept. Since the STFT is a linear transformation, separability guarantees that many pointwise operations, such as energy (squared magnitude), argument, and cross-spectral products behave linearly. For example, if $F(\omega, T)$ is separable at point (ω_0, T_0), and F_n is the dominant component at that point, then energy is linear:

$$\begin{aligned}
&\left| F\left(\omega_0 + \frac{\epsilon}{2}, T_0\right) F^*\left(\omega_0 - \frac{\epsilon}{2}, T_0\right) \right|^2 \\
&= \left| \sum_m \left[F_m\left(\omega_0 + \frac{\epsilon}{2}, T_0\right) \right] \sum_m \left[F^*_m\left(\omega_0 - \frac{\epsilon}{2}, T_0\right) \right] \right|^2 \\
&\approx \left| F_n\left(\omega_0 + \frac{\epsilon}{2}, T_0\right) F^*_n\left(\omega_0 - \frac{\epsilon}{2}, T_0\right) \right|^2 \\
&\approx \sum_m \left| F_m\left(\omega_0 + \frac{\epsilon}{2}, T_0\right) F^*_m\left(\omega_0 - \frac{\epsilon}{2}, T_0\right) \right|^2.
\end{aligned} \tag{7}$$

We define the instantaneous frequency (IF), channelized instantaneous frequency (CIF), and local group delay (LGD) functions as

$$\mathrm{IF}_f(t) = \frac{d}{dt} \arg\{f(t)\}, \tag{8}$$

$$\mathrm{CIF}_f(\omega, T) = \frac{\partial}{\partial T} \arg\{F_h(\omega, T)\}, \tag{9}$$

$$\mathrm{LGD}_f(\omega, T) = -\frac{\partial}{\partial \omega} \arg\{F_h(\omega, T)\}, \tag{10}$$

where the analytic representation of $f(t)$ is assumed in the IF computation, and we assume that $\arg\{F_h(\omega, T)\}$ is continuously differentiable.

It should be noted that the group delay and instantaneous frequency functions, introduced more than 30 years ago, were originally defined as derivatives of the phases of the Fourier transform of the signal and the time-domain signal respectively (cf. Refs. 2, 3). In the literature, the LGD and CIF surfaces, as we have defined them, are sometimes also called the group delay and instantaneous frequency. Since we will relate the CIF and LGD surfaces to the IF function, as we have defined them, we require separate definitions for these functions. The LGD is "local" in the sense that it is defined pointwise on the STFT surface. While the CIF is also a local surface function, it "channelized" in the sense that the filterbank representation of the Fourier transform, discussed briefly in the next section, effectively decomposes the signal into narrow-band channels, indexed by ω. For fixed ω_0, $\mathrm{CIF}_f(\omega_0, T)$ is the IF of the signal in the channel indexed by ω_0. This principle is the basis of the cross-power spectral estimation method.[5]

There are many ways to estimate the CIF and LGD surfaces. For instance they may be estimated as phase differences, as, for example, Kay's spectral estimator.[12] Alternatively, STFT phase may be formally differentiated to produce

a closed form representation computed by Oppenheim and Schafer[13] and used extensively by Yegnanarayana *et al.* in spectral estimation and various speech applications (cf. Refs. 14, 15). For our purposes it suffices to estimate them using two intermediate cross-spectral surfaces

$$C_f(\omega,T,\epsilon)=F_h\left(\omega,T+\frac{\epsilon}{2}\right)F_h^*\left(\omega,T-\frac{\epsilon}{2}\right), \tag{11}$$

$$L_f(\omega,T,\epsilon)=F_h\left(\omega+\frac{\epsilon}{2},T\right)F_h^*\left(\omega-\frac{\epsilon}{2},T\right). \tag{12}$$

The spectrogram (squared magnitude of the STFT surface), LGD, CIF, and mixed partial phase derivative surfaces and may then be estimated as

$$|F_h(\omega,T)|^2\approx|C_f(\omega,T,\epsilon)|\approx|L_f(\omega,T,\epsilon)|, \tag{13}$$

$$\mathrm{CIF}_f(\omega,T)\approx\frac{1}{\epsilon}\arg\{C_f(\omega,T,\epsilon)\}, \tag{14}$$

$$\mathrm{LGD}_f(\omega,T)\approx\frac{-1}{\epsilon}\arg\{L_f(\omega,T,\epsilon)\}, \tag{15}$$

$$\begin{aligned}\frac{\partial^2}{\partial\omega\partial T}\arg\{F_h(\omega,T)\}&\approx\frac{1}{\delta\epsilon}\arg\left\{C_f\left(\omega+\frac{\delta}{2},T,\epsilon\right)C_f^*\left(\omega-\frac{\delta}{2},T,\epsilon\right)\right\},\\&=\frac{1}{\delta\epsilon}\arg\left\{L_f\left(\omega,T+\frac{\epsilon}{2},\delta\right)L_f^*\left(\omega,T-\frac{\epsilon}{2},\delta\right)\right\}\\&\approx\frac{\partial^2}{\partial T\partial\omega}\arg\{F_h(\omega,T)\},\end{aligned} \tag{16}$$

for small values of ϵ and δ. If the surface is separable, locally the CIF, LGD, and mixed partial phase derivatives represent the derivatives of the locally dominant signal component.

If the derivatives are estimated directly as the differences of the undelayed and delayed $\arg\{F_h(\omega,T)\}$ surface, for example,

$$\frac{d}{dt}\arg\{F_h(\omega,T)\}\approx\frac{1}{\epsilon}\left[\arg\left\{F_h\left(\omega,T+\frac{\epsilon}{2}\right)\right\}-\arg\left\{F_h\left(\omega,T-\frac{\epsilon}{2}\right)\right\}\right], \tag{17}$$

the surface must be properly phase-unwrapped to insure continuity of the argument function—a process which is not very reliable in noise. This phase-unwrapping must be performed for the derivative computed this way, even if the STFT surface is a continuous complex valued function of ω and T. The proof of this is trivial, since the argument function is discontinuous, if we assume its principle value range to be $(-\pi,\pi]$. That is, if the argument of a signal increases beyond π, its principal value has a -2π jump discontinuity, which may, in principle, be removed by phase-unwrapping. The phase derivatives may, however, be estimated from cross-spectral surfaces, without the need for phase-unwrapping. If the STFT surface is smooth, there is no need for phase-unwrapping, and there are no discontinuity in the estimated $\mathrm{CIF}_f(\omega,T,\epsilon)$, $\mathrm{LGD}_f(\omega,T,\epsilon)$, and mixed partial surfaces, for small delay values, ϵ and δ. This assertion is proved by simply noting that

$$\arg\{e^{-i\zeta}e^{i(n2\pi+\zeta+\nu)}\}\equiv\nu \tag{18}$$

for small values of ν and all values of ζ and all integers n. Since the right hand side of Eq. (18) is not dependent on n, it follows that there is no need to phase-unwrap the phase derivatives estimated as cross-spectra.

As already noted, in addition to having the STFT phase derivatives represented as their arguments, the surfaces $C_f(\omega,T,\epsilon)$ and $L_f(\omega,T,\epsilon)$ have essentially the STFT power encoded as their magnitudes. This is extremely convenient since the average CIF and LGD may be estimated as the arguments of the means of complex-valued distributions

$$\omega_C=\frac{1}{\epsilon}\arg\left\{\int_R C_f(\omega,T,\epsilon)\,d\omega\,dT\right\}, \tag{19}$$

$$\omega_L=\frac{1}{\epsilon}\arg\left\{\int_R L_f(\omega,T,\epsilon)\,d\omega\,dT\right\}, \tag{20}$$

where R is a region of the TF plane over which the signal CIF or LGD is assumed to be nearly constant.

In Eqs. (19), (20) we are interested in estimating an average phase of signal components which are assumed to be locally stationary in some region of the TF plane. We could simply average the phase, but that would equally weight the contribution of each surface element to the estimated average phase. If we assume that high surface energy at a point (ω_0,T_0) indicates high local SNR at that point and we assume that the noise phase is random, we see that weighting the surface phase contribution by the surface energy in Eqs. (19),(20) de-emphasizes the noise contribution and results in a average phase estimate which better reflects the signal contribution. The average phase computed this way has been shown to converge rapidly and provide an extremely accurate estimate of the signal frequency at very low SNR.[6,16] Variants of this process have been used to accurately estimate pitch,[10] blindly recover the carrier frequency of single sideband suppressed carrier (SSBSC) signals from the pitch harmonics of the transmitted speech signal[8] and provide a correlation function which suppresses all harmonics and subharmonics of a signal, except for the fundamental.[4,5]

A. Interpretation of Phase Derivatives and Re-mapping

The CIF, LGD, and mixed partial derivative surfaces have simple interpretations. To see these, we note first that the STFT surface may be represented as a filter bank. It is well known (cf. Ref. 17) that, for fixed ω_0, the STFT $F_h(\omega_0,T)$ is the original signal filtered by a filter whose impulse response is

$$h_{\omega_0}=h(t)e^{i\omega_0 t}, \tag{21}$$

i.e.,

$$F_h(\omega_0,T)=f*h_{\omega_0}, \tag{22}$$

where "*" represents convolution. If the filter frequency response

$$H_{\omega_0}(\omega)=\int_{-\infty}^{\infty} h_{\omega_0}(t)e^{-i\omega t}\,dt \tag{23}$$

has its energy concentrated in the positive spectrum, that is, if $|H_{\omega_0}(\omega)|^2<\epsilon$, for $\omega<0$ and small ϵ, then

$$\left|\int_{-\infty}^{\infty} F_h(\omega,T)e^{-i\omega T}\,dT\right|^2=|F(\omega)H_{\omega_0}|^2<|F(\omega)|^2\epsilon. \tag{24}$$

If ϵ is sufficiently small, it follows that $F_h(\omega_0,T)$ is essentially the filtered *ANALYTIC* representation of the signal, even if the input signal is real. The STFT, therefore, effectively distributes the analytic signal in time and frequency. This concept of the STFT as a filterbank with analytic output is fundamental to the concept of re-mapping, and will be used extensively in developing the re-mapping principles and in interpreting re-mapped STFTs of speech.

We again assume the AM/FM model of Eq. (5) and assume that there is one signal component f_n [i.e., $F_n(\omega,T)$] which is dominant near some separable point (ω_0,T_0). By separability, f_n must be dominant in $(\omega_0\pm\epsilon,T_0\pm\delta)$, for small ϵ and δ. From the filterbank representation of the STFT, Eq. (22), we may therefore interpret the CIF as the instantaneous frequency of the filtered analytic representation of f_n dominant at that point. The instantaneous frequency is an instantaneous estimate of the rate of change of phase with respect to time, and therefore represents a re-estimation of the frequency of the dominant signal component observed at point (ω_0,T_0). Similarly, the local group delay, $\mathrm{LGD}_f(\omega_0,T_0)$, may be interpreted as the average time of $F_h(\omega_0,T_0)$.[11] That is, the local group delay is the time of the coordinate (ω_0,T_0), computed by averaging over the integration time of the Fourier transform. Assuming separability, the local group delay represents an estimate of the timing error of the dominant signal component observed at point (ω_0,T_0).

For each point (ω_0,T_0), the CIF and LGD effectively provide a re-estimation of the time and frequency of the signal component which resulted in the STFT response at (ω_0,T_0). We would therefore like to re-map the STFT surface by relocating the surface components from their observed locations to their re-estimated locations. This re-mapping has the representation

$$(\omega,T)\rightarrow[\mathrm{CIF}_f(\omega,T),T+\mathrm{LGD}_f(\omega,T)]. \tag{25}$$

The re-mapping concept is quite simple. If we compute the STFT of a single FM component, the STFT response to that signal is distributed in time and frequency. We assume, however, that the value of the STFT at each point (ω_0,T_0) is the response to an event which occurred at a single point in time and frequency. For the phase evolution of the STFT surface to be consistent, an STFT response observed at (ω_0,T_0) must be the response to an event which occurred at the time and frequency represented by the re-mapped point $[\mathrm{CIF}_f(\omega_0,T_0),T_0+\mathrm{LGD}_f(\omega_0,T_0)]$.

In general, the re-mapping is many-to-one, and need not be onto, as we will see in the example of the next section, where we show that the entire STFT surface of any linear-FM signal is re-mapped to a single straight line. The phase ambiguity of the argument function plays no role in this many-to-one relationship. It is many-to-one since more than one STFT component may represent the same FM signal component.

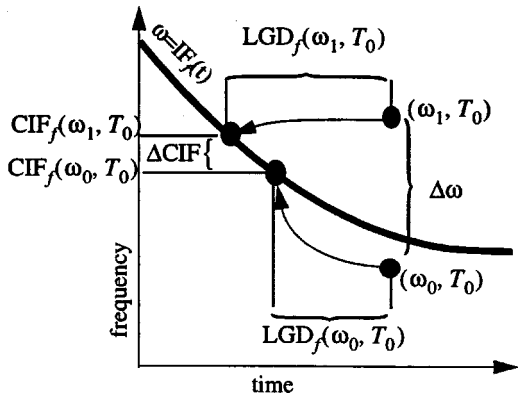


FIG. 1. Mixed partial derivative test for stationarity. If $|\Delta\mathrm{CIF}|>|\Delta\omega|$, the re-mapped components are more dense in frequency than their pre-images on the original surface, the dominant signal component is approximately stationary.

Finally, we supply an interpretation of the mixed partial phase derivatives of the STFT surface. The mixed partial derivatives may be represented as

$$\frac{\partial^2}{\partial\omega\,\partial T}\arg\{F_h(\omega,T)\}=\frac{\partial}{\partial\omega}\mathrm{CIF}_f(\omega,T), \tag{26}$$

$$\frac{\partial^2}{\partial T\,\partial\omega}\arg\{F_h(\omega,T)\}=-\frac{\partial}{\partial T}\mathrm{LGD}_f(\omega,T). \tag{27}$$

By the the representations Eqs. (14) and (15), the mixed partial derivatives of Eqs. (26) and (27) must be equal. If the mixed partial derivatives are zero in a neighborhood of (ω_0,T_0), by Eq. (26), the re-mapping in this neighborhood reduces to

$$(\omega,T)\rightarrow(\Omega_0,T), \tag{28}$$

for some constant Ω_0. That is the dominant signal component in that neighborhood must be locally stationary. If the mixed partial derivatives equal -1, the rate of change of the LGD is the negative of the rate of change of the time coordinate T, and the re-mapping Eq. (25) reduces to

$$(\omega,T)\rightarrow(\omega,T_0), \tag{29}$$

for some constant T_0, which is the condition for an impulse in time. The mixed partial derivatives, therefore, provide us with a simple test for stationarity and impulsivity of the locally dominant signal component (Fig. 1).

B. Re-mapping of linear-FM surfaces

We now show that the STFT of an FM signal may be re-mapped to the curve represented by its functional FM representation. We derive the re-mapped STFT surface for a general linear-FM (chirp) signal and show that, for a chirp, the entire STFT surface theoretically re-maps to the functional FM form of the chirp

$$\omega(t)=\frac{d}{dt}\arg\{f(t)\}. \tag{30}$$

This argument may then be extended to general FM signals which may be approximated by piecewise linear-FM signals. In effect this is a proof that re-mapping corrects for the blurring in both time and frequency of the STFT surface.

Let $f(t)$ be a chirp and $h(-t)$ a real analysis window, not identically zero

$$f(t)=e^{i(\alpha t^2+\beta t+\gamma)}, \tag{31}$$

$$f(t+T)=e^{i(\alpha T^2+\beta T+\gamma)}e^{i(\alpha t^2+(\beta+2\alpha T)t)}. \tag{32}$$

The CIF and LGD surfaces may be computed from the argument of the STFT as

$$\arg\{F_h(\omega,T)\}=\alpha T^2+\beta T+\gamma + \frac{1}{2i}\left(\log\int_{-\infty}^{\infty}e^{i(\alpha t^2+(\beta+2\alpha T-\omega)t)}h(-t)dt\right) - \log\int_{-\infty}^{\infty}e^{-i(\alpha t^2+(\beta+2\alpha T-\omega)t)}h(-t)dt\Bigg), \tag{33}$$

$$\mathrm{CIF}_f(\omega,T)=2\alpha T+\beta+\alpha\left(\frac{\int_{-\infty}^{\infty}te^{i(\alpha t^2+(\beta+2\alpha T-\omega)t)}h(-t)dt}{\int_{-\infty}^{\infty}e^{i(\alpha t^2+(\beta+2\alpha T-\omega)t)}h(-t)dt} + \frac{\int_{-\infty}^{\infty}te^{-i(\alpha t^2+(\beta+2\alpha T-\omega)t)}h(-t)dt}{\int_{-\infty}^{\infty}e^{-i(\alpha t^2+(\beta+2\alpha T-\omega)t)}h(-t)dt}\right), \tag{34}$$

$$\mathrm{LGD}_f(\omega,T)=\frac{1}{2}\left(\frac{\int_{-\infty}^{\infty}te^{i(\alpha t^2+(\beta+2\alpha T-\omega)t)}h(-t)dt}{\int_{-\infty}^{\infty}e^{i(\alpha t^2+(\beta+2\alpha T-\omega)t)}h(-t)dt} + \frac{\int_{-\infty}^{\infty}te^{-i(\alpha t^2+(\beta+2\alpha T-\omega)t)}h(-t)dt}{\int_{-\infty}^{\infty}e^{-i(\alpha t^2+(\beta+2\alpha T-\omega)t)}h(-t)dt}\right). \tag{35}$$

For the chirp of Eq. (32), Eq. (34) and Eq. (35) may be combined, resulting in the following relationship between the CIF and LGD surfaces, which is valid for every point (ω,T):

$$\mathrm{CIF}_f(\omega,T)=2\alpha[T+\mathrm{LGD}_f(\omega,T)]+\beta. \tag{36}$$

The LGD represents the estimated relative time delay of the signal at (ω,T), and the CIF represents the absolute re-estimated frequency. Although the signal is observed at (ω,T), the rate at which the STFT phase is rotating as a function of frequency and time, respectively, is consistent only with the time and frequency of an event which occurred at

$$[\mathrm{CIF}_f(\omega,T),T+\mathrm{LGD}_f(\omega,T)]. \tag{37}$$

If we formally define a function, t of two variables ω and T by

$$t(\omega,T)=T+\mathrm{LGD}_f(\omega,T), \tag{38}$$

Eq. (37) formally reduces to

$$\mathrm{CIF}_f(\omega,T)=2\alpha t(\omega,T)+\beta. \tag{39}$$

The functional representation of the chirp IF

$$\mathrm{IF}_f(t)=\frac{d}{dt}\arg\{f(t)\}=2\alpha t+\beta \tag{40}$$

has the same identical form as Eq. (37) and is valid for all t. It follows that the re-mapping process re-maps every point on the STFT surface to the line Eq. (40) representing the functional form of the chirp.

C. Limiting cases of linear-FM surfaces

Equation (39) is theoretically valid for any chirp, and any real windowing function $h(-t)$ not identically zero. There are two limiting cases. The first of these is a sine wave, in which case $\alpha=0$. In this case, by evaluating Eq. (35) and Eq. (34), it is easily verified that

$$\mathrm{CIF}_f(\omega,T)\equiv\beta, \quad \mathrm{LGD}_f(\omega,T)\equiv 0. \tag{41}$$

The second limiting case is an impulse in time. This is the limiting case as $\alpha\to\infty$. Since the re-mapping properties derived above are independent of α, we have, from Eq. (39), the limiting relationship

$$\lim_{\alpha\to\infty}[T+\mathrm{LGD}_f(\omega,T)]\equiv 0. \tag{42}$$

In other words, in the case of a STFT surface representing an impulse in time, re-mapping collapses the entire STFT surface to the line $T\equiv$*constant*. The question is, what is the constant. We could analytically derive the constant by formally computing the LGD of an impulse. We will, however, provide this simpler derivation. Assume, without loss of generality, that $\beta=0$. Each chirp in the class then passes through DC (i.e., $\omega=0$) at time $t=0$. For each α, the STFT surface is re-mapped to the line of Eq. (37). In the limit, the surface must be re-mapped to the line $T=0$. Therefore, for the STFT of an impulse, $\mathrm{LGD}_f(\omega_0,T_0)$ represents the time difference between the time of occurrence of the impulse and the time of the observation, T_0.

Finally, let us assume that the chirp is truncated in time, with a time support interval $[t_1,t_2]$. We state, without proof, that we may expect the STFT surface components near the truncated chirp to be re-mapped to the truncated chirp. Therefore, if we can approximate an FM signal as the limit of piecewise linear-FM signals, we should expect that re-mapping the STFT surface of the FM signal should result in the surface components being mapped to the functional FM representation of the signal Eq. (30). There is one caveat. For such a re-mapping to be completely correct for a component of $F_h(\omega_0,T_0)$, the FM modulating function, Eq. (30), must be invertible within the effective time interval over which the Fourier transform was computed. A single surface component can only be re-mapped to one point, and, if the representation is not invertible, there is a TF ambiguity, which cannot be completely resolved. This ambiguity is unrelated to the $2n\pi$ phase ambiguity of the argument function. The ambiguity is a function only of the local invertibility of the IF representation of the signal component.

III. RE-MAPPING SIMULATIONS

In the previous two subsections, we have proven that the entire re-mapped image of a STFT of a linear-FM signal is the line in the time-frequency plane which is the FM representation of the signal. We will demonstrate that surface re-

mapping can accurately estimate the instantaneous frequency of FM modulated signals. There are many techniques used to estimate such signals. For single-component analytic signals $f(t)$, the instantaneous frequency can be estimated exactly as $\mathrm{IF}_f(t)$. For single-component real signals, the analytic signal may first be estimated by using the Hilbert transform to estimate the imaginary part of the analytic signal. The accuracy of this method is limited by the quality of the Hilbert filter. Maragos *et al.* proposed using the Teager operator, which they demonstrated performed better than the Hilbert method.[18] Each of these methods may perform well for single-component FM modulated signals, but they cannot be used on multicomponent signals, since they estimate the instantaneous frequency of the ensemble signal, which generally does not reflect the frequency of any of the components.[19] Quatieri *et al.*[17] and others have proposed a balanced frequency demodulation method based on the comparison signal energy in each of two filters. They demonstrated that the method performs well, but the performance was only demonstrated for single component FM signals at fairly high SNR. Beam-forming techniques, such as the MUSIC algorithm perform quite well in estimating multiple sine waves, but assume that the number of sine waves is known.[20,21] Each of the methods mentioned here addresses only the problem of frequency estimation and not the joint estimation of the signal in time and frequency. Spectral estimation algorithms, such as MUSIC and balanced frequency demodulation, which are based on filters, or the Fourier transform, do not consider the possibility that a signal component may be more localized in time than the analysis interval of the transform or the impulse response of the filter. By contrast, the re-mapping proposed here does not require modeling assumptions, is insensitive to peak picking errors, and may be used effectively on multicomponent nonstationary signals. It is a joint time-frequency estimation technique, which can estimate signal components jointly in time and frequency to more accuracy than the Heisenberg limit. This would seem to be a contradiction, were it not for the fact that we assume an AM/FM modulation model and separability, which are not assumed in the Heisenberg uncertainty principle.

To this point, we have assumed a continuous signal and a continuous and infinite Fourier transform. In practice, these assumptions are almost never valid. In most applications, we are forced to process digital signals using short, finite length analysis windows. In all of the simulations and applications to speech which follow, we will assume the more realistic case of sampled nonstationary signals, which must be estimated from extremely short data sets. In assuming a discrete model, the TF representation is quantized in both time and frequency. On the STFT surface, we cannot isolate an event in time and frequency with more precision than the time and frequency quantization error of the surface. If we consider a discrete Fourier transform (DFT) computed on an analysis window which is n samples long, the DFT partitions the spectrum from DC to the sample rate of the signal into n spectral bands of equal width (cf. Ref. 13). The value of the DFT representing the nth spectral band we call the nth Fourier coefficient. Since each coefficient represents a nonoverlapping spectral band, the precision obtained in representing frequency by the index of the spectral band in which it is detected is equal to the width of the band. If we estimate the frequency of a sine wave by peak-picking the DFT (representing the frequency of the sine wave as the index of the DFT coefficient having maximum magnitude), the error in the estimate is the difference between the actual signal frequency and the frequency represented by the coefficient with maximal energy. Assuming we pick the DFT coefficient representing the frequency which is nearest the actual signal frequency, we still may have a maximum peak pick error (MPPE) as large as half the frequency precision of the DFT or

$$\mathrm{MPPE} = \pm \frac{1}{2n} (\text{sample rate}). \tag{43}$$

For small analysis windows (i.e., small n), the peak-pick error can be quite large.

We demonstrate the accuracy of re-mapping and the principles of the previous section in Figs. 2–4. Figure 2 represents the re-mapping of sine waves and impulses. Figure 3 represents the re-mapping of linear-FM (chirp) signals, and Fig. 4 represents the re-mapping of the surface of a quadratic FM signal of the form

$$f(t) = e^{i\alpha t^3 + \beta t^2 + \gamma t + \phi}. \tag{44}$$

In Figs. 2(a)–(c), two sine waves of different frequencies were added together, and the resulting signal was truncated to 128 samples. The signal was windowed with a rectangular window and zero filled to 256 samples. $\mathrm{CIF}_f(\omega, T_0)$ and $\mathrm{LGD}_f(\omega, T_0)$ were computed for a single value of T_0 (i.e., from a single DFT). In Fig. 2(a), the power spectrum (dB)

$$\mathrm{PS}(\omega) = 10 \log_{10}(|F(\omega)|^2)$$

of the signal is plotted as a continuous curve, and the result of re-mapping each individual power-spectral coefficient is displayed as a dot. Figure 2(b) is an expansion of a portion of the spectrum represented in Fig. 2(a). In Fig. 2(c), the result of re-mapping the power spectrum by the group delay function is displayed. In this figure, each point of the power spectrum is re-mapped to a point whose X-coordinate is the local group delay and whose Y-coordinate is the power spectrum (dB) of Fig. 2(a). In Fig. 2(d), the STFT surface of a single impulse was computed, using a rectangular window of length 128 and a transform size of 512. The magnitude of the STFT surface is displayed as a gray-scale image. The surface was then re-mapped, and the result of the re-mapping is superimposed on the original STFT display. Under the re-mapping, the entire STFT surface collapsed to a single vertical line (white), which represents the precise time of the synthesized impulse, with no measurable error. The significance of Fig. 1 is that, while the DFT and the STFT represent distributions of the signal in time and frequency, the re-mapped STFT surface of a sine wave is an impulse in frequency, and the re-mapped STFT surface of a time impulse is an impulse in time.

The spectrum of the signal consisting of the sum of the two tones of Fig. 2 was estimated as the normal Fourier

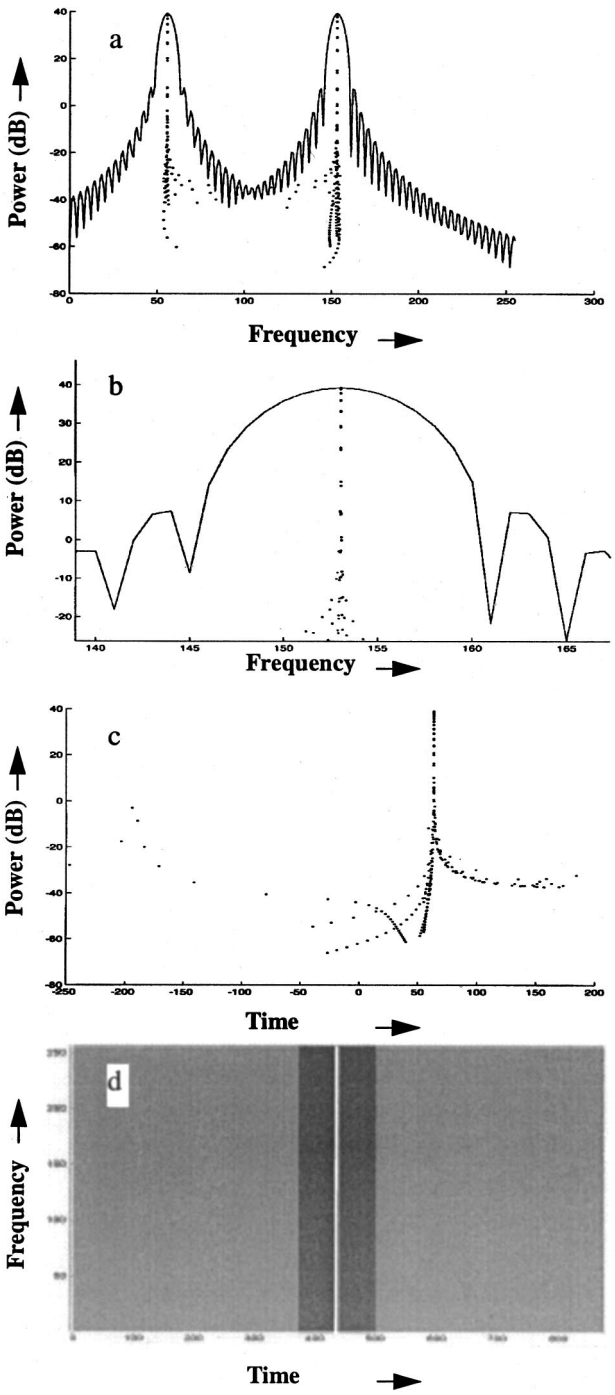


FIG. 2. Re-mapping of sine waves and an impulse. (a) A single 512 point power spectrum (dB) of 128 samples of the sum of two sine waves. Solid line: power spectrum, dotted line: re-mapped spectrum; (b) expansion of a portion of (a); (c) group delay (*X*-axis) versus log power (dB) of the data in (a); (d) spectrogram of an impulse and re-mapped spectrogram. The vertical white dotted line is the locus of the entire re-mapped surface.

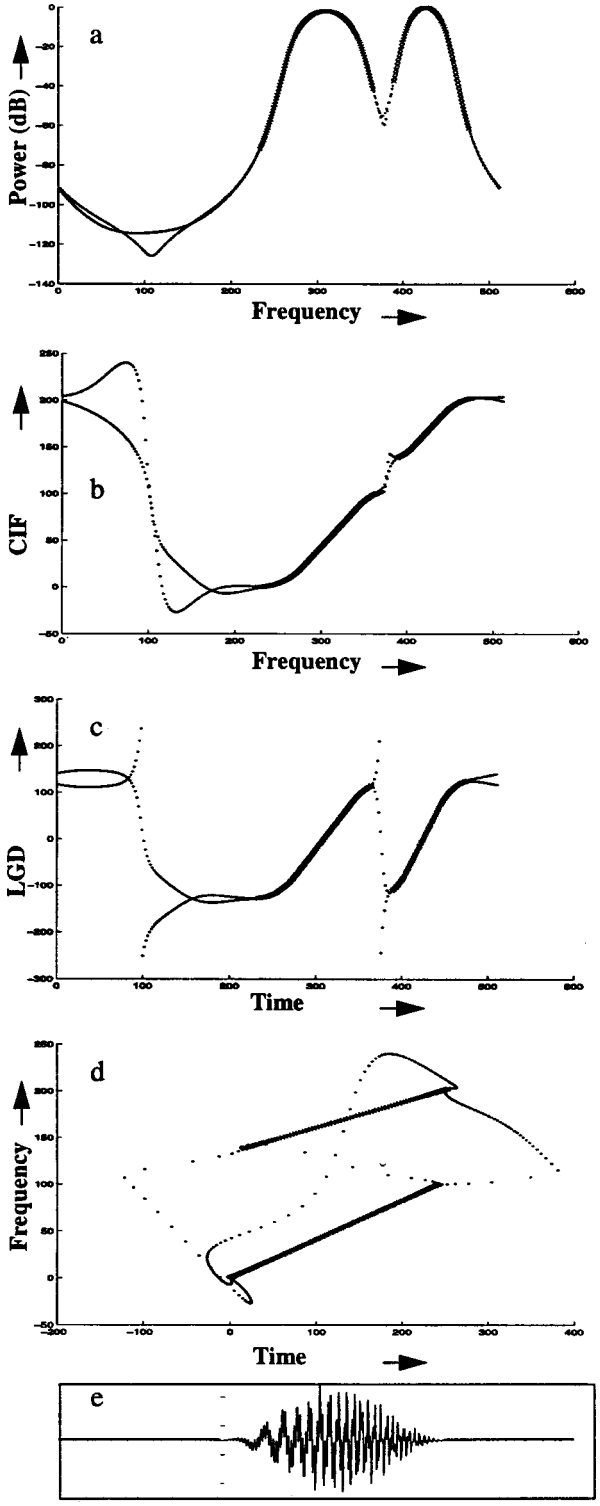


FIG. 3. The sum of two chirps (linear-FM modulated signals) with different chirp rates re-mapped from a single 512 point Fourier transform. The data consist of 256 samples of signal Hanning windowed and zero-filled to 512 samples. (a) Power spectrum (dB); (b) CIF; (c) LGD; (d) re-mapped Fourier transform; (e) original signal. In each graph, highlighted regions by represent regions of nearly perfect reconstruction.

power spectrum (PS), the CIF re-mapped power spectrum (CS) and the spectrum estimated using the MUSIC algorithm in the MATLAB SIGNAL PROCESSING TOOLKIT. A comparison of the three methods was made at several SNRs, and the results of the comparison of the methods for clean signals and at 0 dB are presented in Tables I and II. As represented in Table I, the CS estimates in this example are 2500 times more accurate than the frequency precision of the DFT and 1000 times more accurate than the estimates made by peak-picking the DFT.

The peak error in each case was computed as the frequency difference between the actual signal frequency and the frequency of the energy peak in each of the spectral

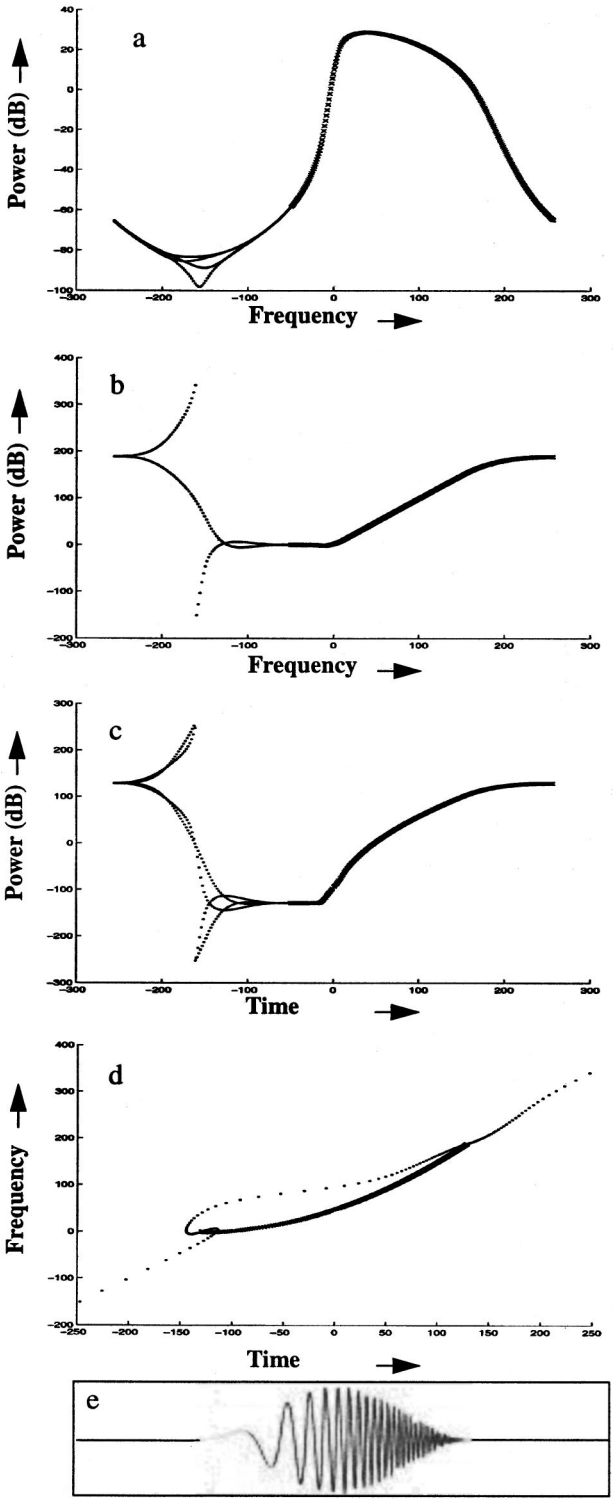


FIG. 4. A quadratic FM modulated signal re-mapped from a single 512 point Fourier transform. The data consist of 256 samples of signal Hanning windowed and zero-filled to 512 samples. (a) Power spectrum, (dB); (b) CIF; (c) LGD; (d) re-mapped Fourier transform; (e) original signal. In each graph, highlighted regions by represent regions of nearly perfect reconstruction.

representations. For the PS and MUSIC algorithms, this error is dependent on the frequency quantization of the DFT, and the value 0.444 represents the quantization error for this example. For the CS algorithm, the estimated frequency is the CIF re-mapped frequency and is therefore not quantized. The bandwidth at several energy levels was computed by measuring the width of the spectral bulge produced by each method at the stated power level. For example, the 5 dB bandwidth of the CS algorithm is the width of the re-mapped power spectrum measured at a spectral power level which is 5 dB below the peak spectral power of the signal. For the clean signal, the spectral distributions representing each of the sine waves was identical, so only one set of numbers is presented in Table I. At 0 dB SNR, the distributions were different, so a set of estimates is presented for each of the two sinusoids. At 0 dB SNR, the MUSIC algorithm could not detect the first tone, and neither tone could be detected by any of the algorithms in the spectral components whose power was less than 20 dB of the peak power. To illustrate this point, the PS and CS representations are depicted in Fig. 2(a). The two tones are detectable at the 70 dB threshold, but not at 80 dB because the null in the spectral envelope between the two bulges does not extend 80 dB below the peak power.

Figure 3 represents the re-mapping of a single DFT of the sum of two chirps with slightly different chirp rates. The data for this simulation consist of 128 samples of signal, Hanning windowed, and zero filled to 512 samples. The DFT, CIF, and LGD (vectors) were estimated for a single value of T_0, and the results are displayed. In Fig. 3(a), the normal power spectrum (dB) is displayed. In Fig. 3(b), the channelized IF of the transform is displayed, with conventional Fourier frequency on the *X*-axis, and CIF re-mapped frequency on the *Y*-axis. Figure 3(c) represents the LGD computed from the single DFT, with conventional Fourier frequency on the *X*-axis and group delay on the *Y*-axis. Figure 3(d) represents the signal spectrum re-mapped by the LGD and CIF functions. In each of the displays in Fig. 3, the entire spectrum is processed and displayed. For each DFT coefficient, its re-mapped frequency is computed and a dot is displayed at that re-mapped frequency. A spectral band around the spectral bulges corresponding to each of the chirped signal components was then selected. For the first bulge, the lower and upper frequency band limits are the points at which the spectral power drops to approximately 75 dB and 45 dB below peak power, respectively. The band limits for the second bulge correspond to approximately 40 dB and 65 dB below peak power, respectively. The data representing each of these bands are highlighted by *X*'s in each of the displays in Fig. 3. It should be noted that the highlighted sections are not linear when re-mapped by the CIF or LGD alone, but these sections are nearly linear when re-mapped jointly by the CIF and LGD. The joint TF errors in the highlighted re-mapped points of Fig. 3 are better than the Heisenberg uncertainty limit, i.e., for this case, $\Delta\omega\Delta T<\frac{1}{2}$, whereas the Heisenberg limit ($\Delta\omega\Delta T>\frac{1}{2}$) is a sharp lower bound for the resolution of the STFT. The extraneous dots on each of these displays represent spectral components outside of the two spectral bands described above and therefore represent spectral components whose power is a minimum of 40 dB below the peak spectral power of the signal.

Figure 4 represents the re-mapping of a single DFT computed from a truncated quadratic FM signal of the form represented by Eq. (44), with a nonzero value of α. The signal was truncated to 256 samples, Hanning windowed,

TABLE I. Estimation of two tones in a clean environment.

Method	Peak error	5 dB BW	10 dB BW	15 dB BW	20 dB BW	60 dB BW
CS	0.000 39	0.000 77	0.0028	0.0028	0.0037	2.64
PS	0.444	8	10	11	13	52
MUSIC	0.444	2	3.5	6	9.75	N/A

and zero filled to 512 samples. The power spectrum, CIF, and LGD functions were computed and displayed as in Fig. 3. The respective lower and upper band limits of the region highlighted by *X*'s are set at approximately 60 dB and 70 dB below peak power. The re-mapping accuracy of the highlighted region is again better than the Heisenberg uncertainty and is nearly identical to that of the chirps of Fig. 3.

IV. AN APPLICATION TO SPEECH

In the previous sections, we presented the concepts of local group delay and channelized instantaneous frequency as local properties of the STFT surface. Our purpose in doing so was to establish a re-mapping of the STFT surface which concentrates the surface energy of multicomponent FM signals around curves functionally representing the individual FM components. We wish to apply the tools developed in the previous sections to the problems of estimating the instantaneous frequencies of the individual speech formants and the group delay functions of both the vocal tract and the transmission channel.

There have been many methods used for estimation of speech components. Many, if not most, of these methods are based on linear prediction[22] or the Fourier transform. Since we have discussed methods based on the Fourier transform earlier, we make a few comments concerning linear prediction methods before continuing. Linear prediction may be accurate, but generally involves solving for the roots of the LPC polynomial, peak-picking the LPC spectrum, or peak-picking the group delay spectrum, as proposed by Yegnanarayana.[23] In addition, for good results, the order of the LPC polynomial must reflect the number of formants present in the spectrum. As was noted by Welling and Ney, peak-picking can result in merged components and spurious peaks, and root finding is only successful for complex-conjugate poles and not real poles.[24] The method presented by Welling and Ney is based on modeling the resonances within bands. While this method may not have the problems associated with peak-picking or polynomial root estimation, it still requires an estimation of the proper number of formants and their approximate bands. As they note, if this model is incorrect, the results can be erroneous. All of these methods may perform well, but they only address the problem of estimation of the signal frequency. Time is estimated only to the precision of the analysis window, or the length of the prediction polynomial. Re-mapping does not require modeling assumptions, is insensitive to peak-picking errors, and can provide joint time-frequency estimates of the multiple signal components with much better frequency accuracy than peak-picking and much better time accuracy than the window or prediction polynomial length, as we have seen in the previous section.

TABLE II. Estimation of two tones in 0 dB noise.

Method	Peak error	5 dB BW	10 dB BW	15 dB BW
CS	0.086 98	1.5	4	11
PS	0.444	7	12	15
MUSIC	0.5561			
CS	0.009 75	0.4	1.4	3
PS	0.444	7.5	11.5	12
MUSIC	4.04	16	25	32

As we have seen, under the AM/FM model of Eq. (5), for each point (ω_0, T_0), the CIF represents an estimate of the instantaneous frequency of the signal component dominant in a neighborhood of (ω_0, T_0). This estimate has been shown to be essentially independent of the windowing function used in computing the STFT and of the AM modulation. The effective time of the ''instantaneous'' estimate, however, is not T_0. The effective time is the time T_0 of the observation, corrected by the local group delay function $\mathrm{LGD}_f(\omega_0, T_0)$.

For the STFT surface of an impulse, the LGD provides a unique estimate of the time of the impulse, at each frequency. In a transmission channel, it is frequently the case that some frequencies are delayed more than others in transmission. The channel delay as a function of frequency is called the channel group delay. Assume that we are observing the STFT surface of a channel which was excited by an impulse. Since the local group delay is a local property of the STFT surface, each point on the LGD surface will respond with its estimate of the time the impulse was generated. If the channel group delay is not constant as a function of frequency, $\mathrm{LGD}_f(\omega, T_0)$ will be different for each value of ω. For each value of ω, the LGD will provide an estimate of the channel delay for that frequency. That delay is

$$\mathrm{LGD}_f(\omega, T) = g(\omega) + T + \text{constant}, \tag{45}$$

where $g(\omega)$ is the channel group delay, and the additive constant is not frequency or time dependent.

A. A single pulse model for voiced speech

We would like to process the speech signal without having to consider the harmonic effects of a nearly periodic excitation function. The harmonic structure of voiced speech consists of narrow spectral components which are exact integer multiples of the fundamental excitation frequency, $F0$. In estimating formants by phase based methods, these harmonics pose a problem, since phase measurements in the nulls between the harmonics are unreliable, as pointed out by Murthy *et al.*[15] To deal with the spectral nulls, they biased their group delay estimate and smoothed the estimated spectrum. We base our analysis on a wide-band representation of the spectrum in which there are no harmonics and therefore no nulls between the harmonics. $F0$ and its harmonics are

observed in the spectrum only if the length of the analysis window used in estimating the spectrum is longer than the period of the excitation function. For analysis windows shorter than the excitation period, there is no observable excitation periodicity, and there is therefore no observed frequency component representing this periodicity. In all of the speech examples presented here, data from the TIMIT database were used. In each case, a 512 point Fourier transform with a Hanning window of length 81 samples, zero filled to 512 samples was used in the analysis. At the 16 kHz sample rate of the TIMIT database, a period of 80 samples represents a frequency of 200 Hz. For an analysis window of length 81, no artifacts resulting from the periodicity of the excitation function will be present in the spectrum if $F0$ is less than 200 Hz. In fact this window performed well for values of $F0$ below about 400 Hz. With a Hanning window of length 81 samples, separability can be attained for formants whose center frequencies are separated by approximately 300 Hz.

We may therefore consider a model of the speech signal in which the vocal tract is excited by a single broadband glottal pulse. Although the pulse itself may be of relatively long duration, the effective excitation time may, in fact, be relatively short. We make no assumptions concerning the shape, location, or duration of the glottal pulse, and we make no assumption concerning the effective duration of the excitation. We do assume that the bandwidth of the excitation function is sufficiently wide that it excites the entire observed speech spectrum. To make the excitation model completely general, we will assume that the vocal tract itself may have a nonconstant (in frequency) group delay. That is, different frequencies may experience excitation at different times. Following excitation, there is a damped nearly steady state resonance, during which no energy is being added to the system. The damped resonance we model as the multicomponent AM/FM model of Eq. (5), where the AM component represents attenuation caused by friction, lip effects, etc. We assume that the STFT surface is separable and that the AM and FM modulating functions do not change too rapidly during resonance. We make no further assumptions concerning the form of the AM and FM modulations of each resonance. During excitation, the system is driven by the excitation function. If we consider a bank of filters responding to the excitation, there is no particular frequency to which the filters are driven, so we should expect each filter to respond at its natural frequency. The phase of the response, however, is determined by the excitation function, the vocal tract and the channel and should therefore be reflected in the observed group delay. During resonance, we may expect each STFT ''filter'' to respond to the frequency of the resonance which is dominant near that filter frequency.

B. An impulse excitation model

Let us first consider the situation at excitation. Assume for the moment the simple model for vocal tract excitation. In the simple model, the excitation function is an impulse in time, which is experienced by all frequencies at the same time. In this case, it was argued by Murthy *et al.* that the time of the impulse can be estimated as the frequency-averaged group delay.[14] Under this assumption they managed to recover sharp excitation pulses from the LP residuals. We would like to extend this result to a more complex model in which the group delay is not constant.

We note first that at excitation we would expect the filters of the STFT to respond at their natural frequencies. That is we assume $\mathrm{CIF}_{f\mathrm{exc}}(\omega,T)=\omega$. We may accommodate the possibility that the effective time of excitation is frequency dependent by modeling this frequency dependence as a group delay in the vocal tract at excitation. If the group delay of the vocal tract, estimated at a time T_0 of excitation, is constant as a function of frequency, the excitation is experienced by all frequencies at the same time. If it is not constant, then the relative time of excitation at different frequencies is represented by the vocal tract group delay function. In addition, there is an unknown channel group delay. The channel group delay is a measure of the amount that each frequency component of the signal is delayed by the channel. By assuming the three conditions: (1) an impulse excitation function, (2) that the filters respond at their natural frequencies, and (3) that the total group delay is the sum of the vocal tract and channel-group-delays, we obtain the model:

$$F_{\mathrm{exc}}(\omega,T)=A(\omega,T)e^{i((T-T_0)\omega-G(\omega)-G_{\mathrm{exc}}(\omega))}, \tag{46}$$

where $A(\omega,T)$ is slowly varying in time and frequency, T_0 is the fixed time we have assigned to the excitation impulse,

$$g(\omega)=\frac{d}{d\omega}G(\omega) \tag{47}$$

is the channel group delay, and

$$g_{\mathrm{exc}}(\omega)=\frac{d}{d\omega}G_{\mathrm{exc}}(\omega) \tag{48}$$

is the group delay of the vocal tract at excitation. In this case, the LGD and CIF are

$$\mathrm{LGD}_{f\mathrm{exc}}(\omega,T)=T_0-T+g(\omega)+g_{\mathrm{exc}}(\omega), \tag{49}$$

$$\mathrm{CIF}_{f\mathrm{exc}}(\omega,T)=\omega. \tag{50}$$

Note that, under this model, we should expect the LGD to re-map the surface components representing the excitation pulse to a single curve

$$T(\omega)=T_0-g(\omega)-g_{\mathrm{exc}}(\omega), \tag{51}$$

where the right hand side is not dependent on time. We have assumed separability in the sense that the surface components corresponding to excitation are assumed to not interfere with the surface components at steady state resonance.

We will demonstrate experimentally that this excitation model is consistent with the observed STFT of voiced speech.

C. Resonance model

Consider a region on the STFT surface near a formant in nearly steady state resonance, damped by the loss of energy in the vocal tract and at the lips. In this case, we assume that there is no energy contribution from the glottal pulse, and that all filters near the formant frequency are pulled to the formant resonant frequency ω_k in the sense that the output of each of these filters reflects the frequency and phase of the

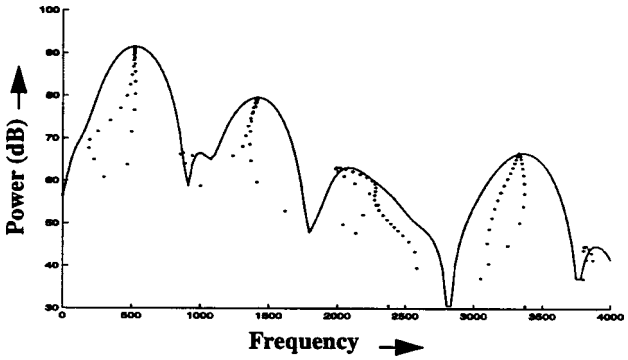


FIG. 5. Power spectrum (dB) and CIF re-mapped power spectrum (dB) of a typical portion of voiced speech. This sample of spectrum corresponds to time $T=220$ in Figs. 7 and 8. Solid line represents power spectrum. Dotted line represents re-mapped spectrum.

formant resonance to which the filter responds. Energy loss in the system results in AM modulation which is observed as a broadening of the observed spectrum.[1] We assume, for modeling purposes only, that the primary factor contributing to the formant bandwidth is the AM modulation and that the FM modulation is comparatively minor. This is an assumption used only for modeling the signal for the purpose of deriving analytic tests for steady state resonance. Obviously there is FM formant modulation resulting from the opening and closing of the glottis and motion of the tongue and lips. If it turns out that the FM is significant, it will be resolved in the re-mapping process. In this resonance case, we may therefore assume that the STFT surface may be locally approximated as

$$F_{\text{res}}(\omega,T)=A(\omega,T)e^{i((T-T_0)\omega_k-G(\omega)-G_{\text{res}}(\omega))}, \tag{52}$$

where the vocal tract group delay during resonance is $g_{\text{res}}=(-\partial/\partial\omega)G_{\text{res}}(\omega)$. The LGD and CIF in this case are

$$\text{LGD}_{f\text{res}}(\omega,T)=g(\omega)+g_{\text{res}}(\omega), \tag{53}$$

$$\text{CIF}_{f\text{res}}(\omega,T)=\omega_k. \tag{54}$$

We have again assumed separability in the sense that we assume that the formants do not significantly interfere with each other, nor do they interfere with the excitation on the STFT surface.

At this point, we provide a simple demonstration of the effectiveness of re-mapping of the speech spectrum. Figure 5 is an example of the re-mapping of the spectrum of speech representing the signal at a time in which the vocal tract is in resonance. This display was computed from a single Fourier transform computed from a portion of a vowel from the TIMIT database. As outlined above, a Hanning window of length 81 samples and a zero-filled 512 point transform were used. The solid line represents the normal power spectrum (dB), and the dotted line represents the re-mapped power spectrum (dB). Clearly the re-mapping process has significantly reduced the effective bandwidth of the speech spectrum. In this example, the bandwidths of the first four formants were estimated at 10 dB below peak power for each formant from the power-spectrum and the re-mapped CIF, and the results are tabulated in Table III. In this example, the indicator functions presented in the next subsection have not been employed. For each formant, the bandwidth reduction is significant. The re-mapping process has been applied to many samples of voiced speech from several databases, and the performance of the re-mapping process in Fig. 5 is representative of the general performance of the process. (The data used in Fig. 5 are the data displayed in Figs. 6–11)

TABLE III. 10 dB bandwidths of first four formants of Fig. 5 measured from power spectrum (PS) and re-mapped CIF (CS).

Method	Formant 1	Formant 2	Formant 3	Formant 4
PS	500	500	600	580
CS	40	80	300	160

In Fig. 6, the first formant region of two excitation periods of speech are displayed. In this figure, the spectrogram is displayed in Figs. 6(a), (c) and the re-mapped surface is displayed in Figs. 6(b), (d). In generating Figs. 6(b), (d), each point of the STFT in the first formant region was re-mapped, and the resulting surface was re-quantized to the same time and frequency quantization as the original spectrogram. The resulting display clearly demonstrates the bandwidth reduction of the re-mapped representation.

D. Mixed partials and indicator functions

In the above discussion, we have modeled the voiced speech signal as an excitation and resonance model, and we have demonstrated by example that the proposed re-mapping of the STFT surface redistributes the surface in a manner which is consistent with our expectations under the model. The mixed partial derivatives provide a convenient test for excitation and resonance. In the excitation case, we may compute the expected mixed phase partial derivatives directly from Eq. (46) as

$$E\left\{\frac{\partial^2}{\partial\omega\,\partial T}\arg\{F_{\text{exc}}(\omega,T)\}\right\}=E\left\{\frac{\partial^2}{\partial T\,\partial\omega}\arg\{F_{\text{exc}}(\omega,T)\}\right\}=1. \tag{55}$$

In the steady state resonance case, from Eq. (52) we may compute the expected mixed phase partial derivatives as

$$E\left\{\frac{\partial^2}{\partial\omega\,\partial T}\arg\{F_{\text{res}}(\omega,T)\}\right\}=E\left\{\frac{\partial^2}{\partial T\,\partial\omega}\arg\{F_{\text{res}}(\omega,T)\}\right\}=0. \tag{56}$$

With these two relationships, we may build indicator functions to test whether any point on the STFT surface is the response to an excitation pulse or a resonance of the vocal tract. The functions

$$I_{\text{exc}}(\omega,T)=\left|1-\frac{\partial^2}{\partial T\,\partial\omega}\arg\{F_h(\omega,T)\}\right|, \tag{57}$$

$$I_{\text{res}}(\omega,T)=\left|\frac{\partial^2}{\partial T\,\partial\omega}\arg\{F_h(\omega,T)\}\right| \tag{58}$$

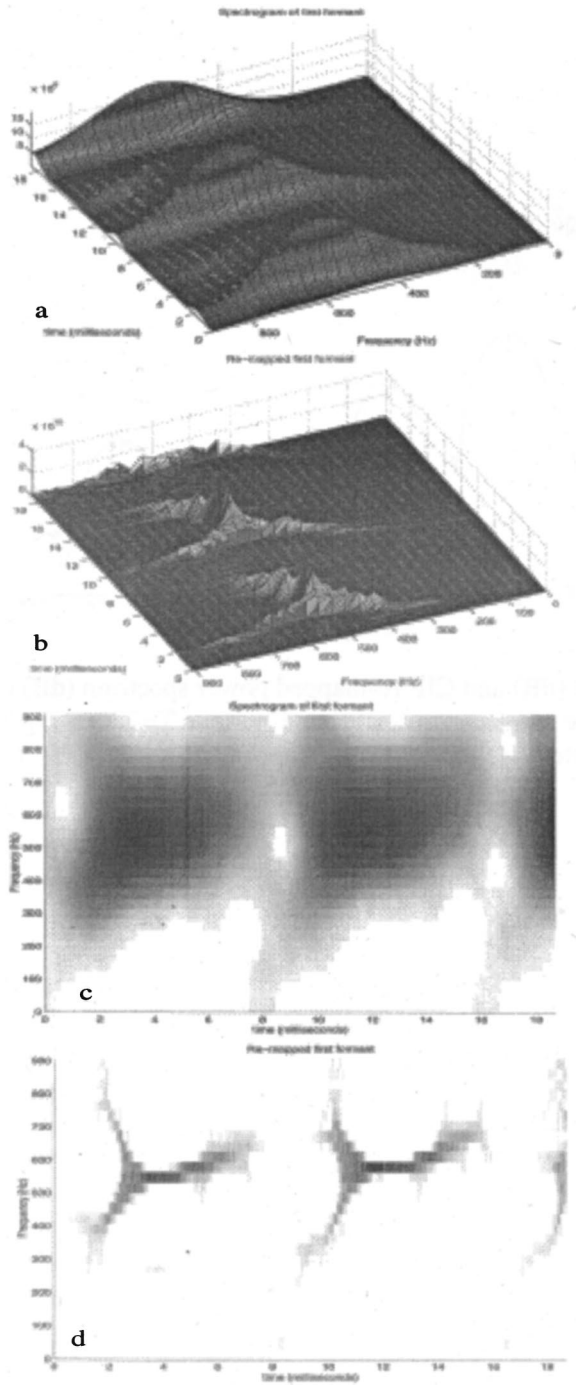


FIG. 6. Two pitch cycles of the first formant of a TIMIT vowel. (a), (c) Spectrogram (power); (b), (d) re-mapped (power). Every sample of (a) was re-mapped to produce (b), (d). The blockiness *n* the displays represents time and frequency quantization of the original STFT. The re-mapped representations have been re-quantized to the TF resolution of the spectrogram.

have expected values zero at excitation and steady state resonance, respectively. If the condition is not approximately met at the point (ω_0, T_0), then the STFT response at that point is not driven by excitation (resonance), and is therefore driven by another process, such as resonance (excitation), noise, or interference. We can therefore discard the points on the STFT surface which are not indicated as the signal condition we are seeking, and therefore improve the processing gain. An important observation is that the indicator functions serve to effectively partition the STFT surface into three surfaces. On one surface, the excitation is dominant, and resonance is effectively removed. On the second surface, resonance is dominant, and excitation is effectively removed. And, on the third surface, artifacts other than resonance and excitation are dominant.

E. Validation of excitation/resonance model

Although the process has been used effectively on data from several different sources, we present a single test sequence, each process of which was applied to the same identical data set. The data selected were from sentence sa1 of speaker mdc10 from the dialect region 3 of the training set of the TIMIT database.

To test the indicator functions, resonance and excitation indicator filters were constructed. The resonance-indicator filter was designed to retain all surface components for which $I_{\text{res}}(\omega,T)<0.5$ and discard all other surface components. The excitation-indicator filter was designed to retain all surface components for which $I_{\text{exc}}(\omega,T)<0.5$ and discard all other surface components. These filters were applied to the data, and the results are presented in Fig. 7. Figure 7(a) represents the spectrogram in dB computed from a short sample of voiced speech from the selected data set. To remove much of the background clutter, all points on the surface not indicated as resonance or excitation have been removed from the display by applying the appropriate indicator filter. Figure 7(b) represents a display of the excitation indicator surface, with the darker areas representing the points indicated as excitation. Figure 7(c) represents the resonance indicator surface, with darker areas representing the points indicated as resonance. Figure 7(d) is a plot of the mixed partial phase derivatives of the surface for values of $T\in[120,130]$. Although there is no way to verify the accuracy of the indicators on real data, the results certainly appear to be reasonable in the sense that most of the areas indicated as excitation are broad spectral bands, relatively localized in time, and most of the areas indicated as resonances represent the frequency bands in which the formants reside, and the majority of the indicated resonance areas immediately follow the indicated excitation in time.

We now apply the indicator filters in re-mapping the STFT surface. Figure 8 represents a series of displays depicting re-mapping in combination with the indicator filters. Figure 8(a) represents the union of the resonance and excitation indicated surfaces re-mapped by the CIF alone. This surface is exactly the spectrogram surface Fig. 7(a) re-mapped using only the frequency re-mapping of the CIF function. Figures 8(b), (c), (d) represent, respectively, the re-mapping of the resonance indicated surface by the CIF alone, the excitation indicated surface by the LGD alone, and the full re-mapping of the union of the excitation and resonance surfaces by both the LGD and CIF functions.

In this display, as in the displays of most other re-mapped surfaces presented, each point of the STFT surface is re-mapped and displayed as a single unmodulated dot. Since the STFT is discrete in time and frequency, there are a finite number of dots displayed. Since the CIF and LGD represent floating-point precision estimates of frequency and timing error, there is no time or frequency quantization in the re-

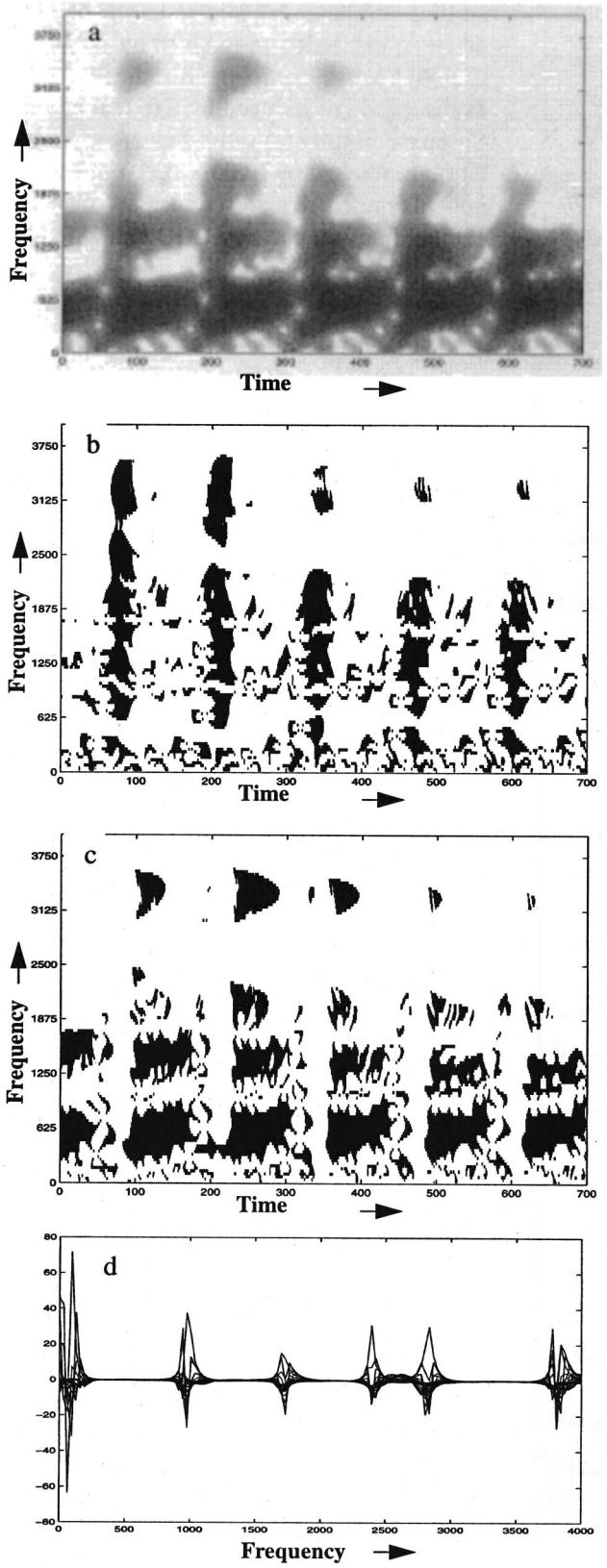


FIG. 7. Spectrogram (dB) and indicator functions computed from the mixed partial phase derivatives. (a) Spectrogram; (b) excitation indicator function; (c) resonance indicator function; (d) mixed partial derivative function corresponding to times $T=[120,130]$ of (a).

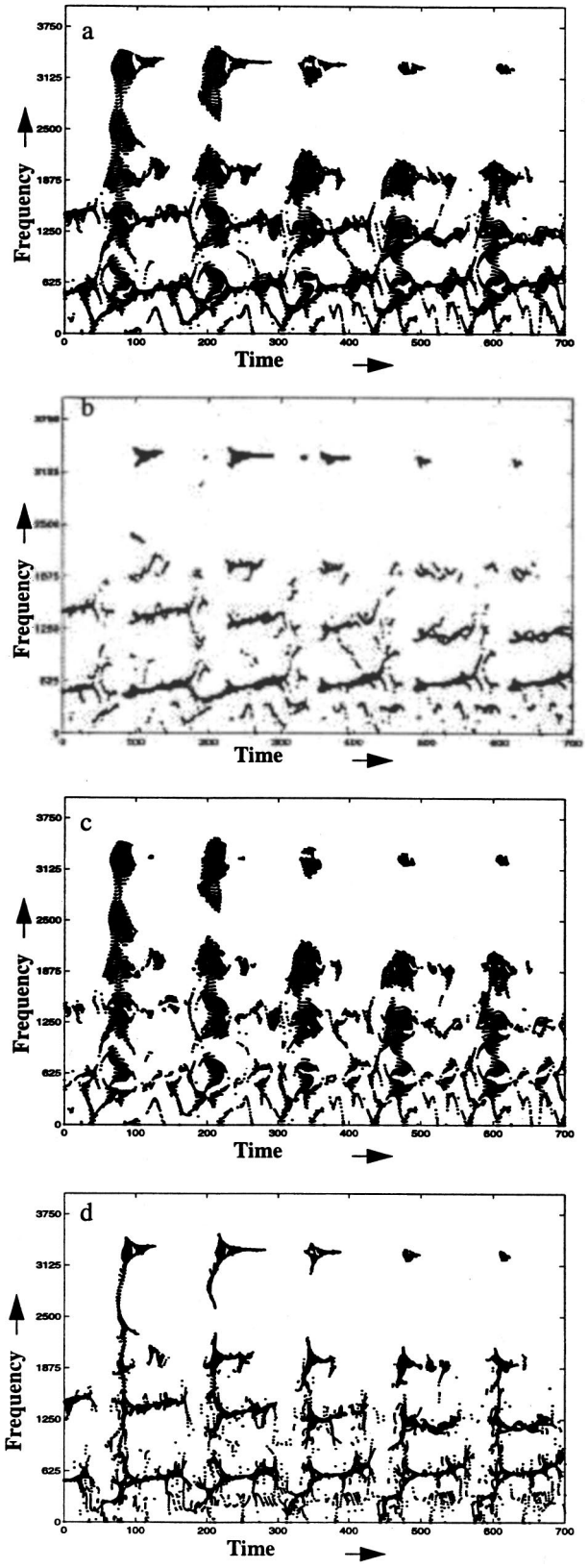


FIG. 8. Re-mapped STFT surface of data represented by Fig. 7. (a) All surface components indicated as resonance or excitation re-mapped by CIF. (b) Surface components indicated as resonance re-mapped by CIF. (c) Surface components indicated as excitation remapped by LGD. (d) All components indicated as resonance or excitation re-mapped by LGD and CIF.

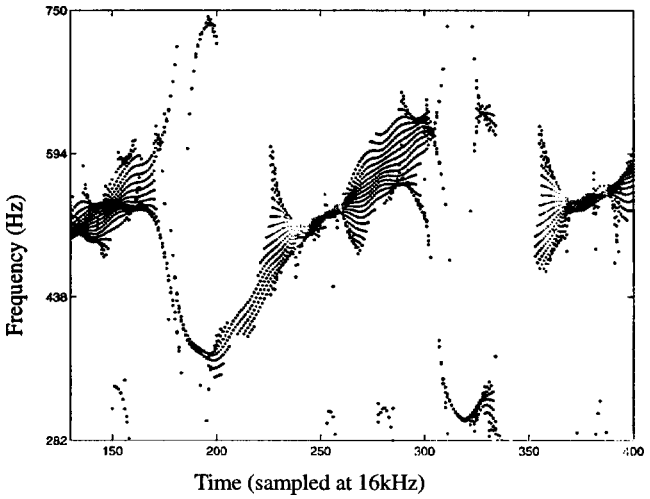


FIG. 9. An expanded portion of the first formant band from Fig. 7(b), showing a nearly linear-FM, with a total excursion of about 250 Hz.

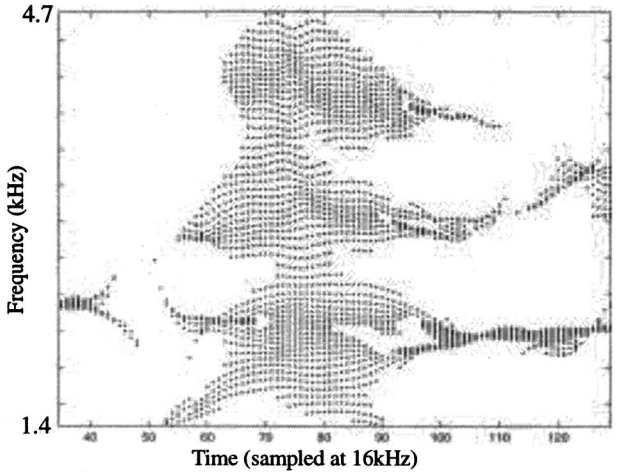


FIG. 10. An expanded view of the CIF re-mapped portion of speech excitation displayed in Fig. 7(c).

mapped surface. However, the re-mapped surface is sparsely represented, except near the signal components. Near the signal components, the representation is dense, demonstrating the convergence of the re-mapping and the attraction of the signal components. In Fig. 6, the re-mapped surface was re-quantized in time and frequency in order to display the surface magnitude. We have presented the re-mapped surface of Fig. 8 as isolated dots in order to show the effects of re-mapping individual points. These displays give a good indication of how the surface is redistributed, but, for a full appreciation of the re-mapping process, one must also look at single trace plot, such as Fig. 5, which represents the single trace of Figs. 7(a), 8(a) at time $T=120$.

The re-mapping of the resonance indicated surface has effectively condensed the STFT surface, to the extent that we can identify, within the pulse intervals, an apparent nearly linear frequency modulation of the formants. A portion of the first formant band is expanded in Fig. 9, showing a nearly linear-FM, with a total excursion of about 250 Hz. The data of Fig. 9 represent the same speaker, approximately the same time within the utterance and the same spectral band as that represented by Fig. 6. In Fig. 6, however, the display was re-quantized in time and frequency. Although formant motion is still apparent in Fig. 6, its motion is less clear, due to the granularity of the display. This intra-pulse formant motion is consistent with the observations of Quatieri *et al.*[25] Equally of note is the observation that the bands, indicated as excitation, re-map by the CIF function to a set of dots which are nearly equally spaced in frequency, as depicted in Fig. 10. This demonstrates that the CIF re-mapping does not significantly change the frequency of the re-mapped spectrum from that of its observed location. This is the condition that the filters of the Fourier transform resonate at their natural frequencies, indicating that, at least for the sample of speech processed in this test, our excitation model is essentially valid. A final, but important observation from these figures is the fact that the indicated excitation condenses to nearly a time-constant, broadband line, which can best be seen in Fig. 8(d) at $T_0=90$. This indicates that, for this particular data set, the group delay of the vocal tract is apparently constant in the sense that all frequencies apparently experience excitation at nearly the identical time. The re-mapping process has been applied to many samples of data from different sources, and, for speech which has not been subjected to a transition channel, the process has consistently indicated a constant vocal tract group delay. This observation will prove crucial in estimating the channel-group-delay function.

F. Blind estimation and equalization of a speech channel

As our final application to speech, we address the problem of blind recovery of the group delay of a transmission channel. In view of the last observation of the previous section, it suffices to estimate the region of the STFT surface which is the response to a single glottal pulse. We may then calculate the re-mapping of that region, and estimate an average group delay by averaging the times of the re-mapped points for each frequency; that is,

$$\gamma(\omega_0)=\langle \mathrm{LGD}_f(\omega,T)-T\rangle, \tag{59}$$

where $\langle \mathrm{LGD}-T\rangle$ represents the ''average'' of $\mathrm{LGD}-T$, and the average is computed for all values of T and ω for which $\mathrm{CIF}_f(\omega,T)=\omega_0$ within the excitation region of a single pulse on the STFT surface. We will not bias the process by specifying how the average must be computed. It may be computed as a mean, angular mean, etc. From our previous remarks, this average group delay represents the sum of the vocal tract group delay and the group delay of the the transmission channel. However, if we can ignore the vocal tract group delay, as indicated by the previous section, the average group delay must be an estimate of the channel group delay.

As a test, we select a portion of data from the TIMIT database. Since this database was collected under essentially studio conditions, we may assume that the channel group delay effects are minimal. This database has a companion database, NTIMIT, which is the TIMIT data subjected to the NYNEX telephone channel. With the TIMIT database, we have nearly undistorted signals, which we can compare to the same signals in the NTIMIT database. Since the two sets of

signals differ only in their channels, we can estimate the channel group delay and phase equalize the NTIMIT channel. By comparing the properties of the equalized signal to the original signal, we may determine whether the channel was properly estimated. We select a portion of NTIMIT data corresponding to the previously selected TIMIT data. We calculate the re-mapped surface of the NTIMIT data indicated as excitation, and we calculate the average group delay for one excitation region. We may validate our estimated group delay by shifting each re-mapped point of the NTIMIT STFT surface by a time equal to $-\gamma(\omega_0)$, where ω_0 is the frequency of the re-mapped point. This process will re-align the excitation region to the broad spectral band of an impulse which may be observed by all frequencies at the same time.

In Fig. 11(a), the re-mapped NTIMIT data are displayed, with a dashed reference line, which highlights the estimated group delay of the signal. Figure 11(b) is a plot of the estimated group delay computed for each value of ω_0 in the spectral band $0<\omega_0<2500$ Hz. For each value of ω_0, $\mathrm{LGD}_f(\omega_0, T_0)$ was computed for ten consecutive values of T_0 selected near the center of the excitation indicated surface for that value of ω_0. The LGD for these points is displayed in Fig. 11(b) as continuous lines, with the estimated group delay plotted as a solid gray line. The gray line represents the approximate angular mean of the LGD function. The reference line in Fig. 11(a) is exactly the same estimated group delay displayed in Fig. 11(b). The boxes in Fig. 11(b) serve to highlight the approximate frequency bands of the first four formants. In Fig. 11(c), the NTIMIT data phase equalized by the estimated group delay function is displayed, with a dashed reference line at the approximate time of the equalized excitation pulse. Figure 11(d) represents a sample of TIMIT data corresponding to approximately the data selected from the NTIMIT signal. The alignment may not be exact, since the exact timing of the two signals is not precisely known. However, the group delay characteristics of the TIMIT signal showed no apparent change in a large time interval around that displayed.

V. PHASE AND HEARING

We wish now to make a brief point concerning phase and hearing. Although the human auditory process is not fully understood, there have been many experiments which have resulted in providing us with a good understanding of much of the process. While there is not enough experimental evidence to support (or refute) a claim that the high performance of the human ear is phase based, we can demonstrate from the literature that spectral phase is used in the auditory process and that some of the performance of humans in distinguishing sounds is consistent with the performance of the phase-based processes presented here. Note that we are not claiming that the ear performs any process similar to the re-mapping process discussed above. We merely argue that spectral phase is used in the auditory process and that the frequency resolution and accuracy of the human ear can be simulated by the methods presented here.

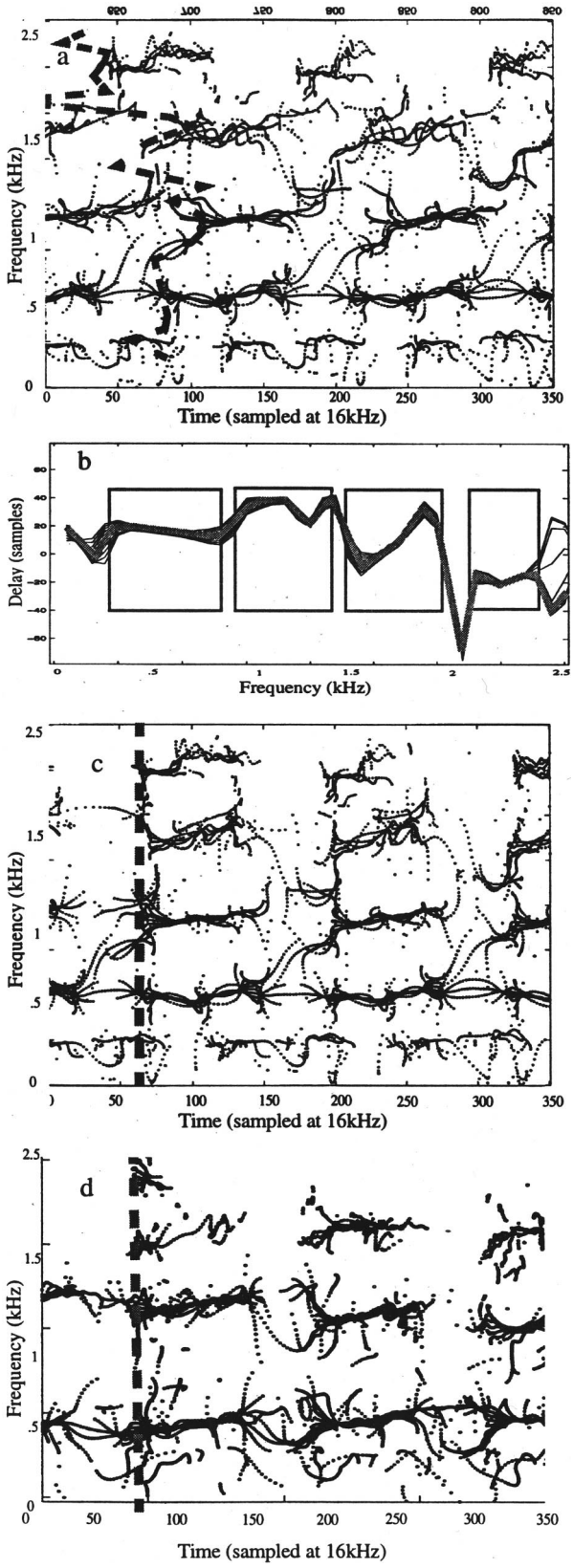


FIG. 11. Phase equalization of a speech signal. (a) Original NTIMIT signal (re-mapped); (b) LGD and estimated group delay, with approximate formant bands indicated by boxes; (c) phase equalized NTIMIT signal (re-mapped); (d) re-mapped TIMIT signal approximately aligned in time with the corresponding equalized NTIMIT signal.

The concept of spectral phase in hearing is not new. It in fact goes back to 1907 when Lord Raleigh published the evidence that the ear uses spectral phase to determine the locations of sounds.[26,27] The experiment involved the use of two tuning forks at slightly different frequencies. When the sounds of the two tuning forks are presented to different ears, the apparent position of the sound moves from left to right and back. The experiment was repeated by von Békésy[28] in an experiment in which he used thin tubes as acoustic waveguides to transmit tones from a common source to the ears. By varying the length of one of the tubes, the relative phase difference of the tone at the two ears could be changed without changing the relative amplitude. Again the apparent direction of the sound changed as the relative phase of the tones at the two ears changed. This phenomenon was observed to hold for frequencies below about 1500 Hz. Although this is a binaural example, it demonstrates that phase is preserved in the auditory process and phase is at least partially utilized in that process.

One might ask what the ear's mechanism is for processing phase. The answer is quite simple. As sound vibrations are transferred to the basilar membrane, a traveling, not a standing, wave is induced on the basilar membrane.[29,28] The significance is that each point of the basilar membrane vibrates up and down as the wave propagates. In effect the basilar membrane functions as a dispersive delay line, in which each position along its length is responsive to a frequency band around a particular resonance frequency. If we excite the basilar membrane with a sine wave and consider the vibration observed at a particular location on the membrane in the response region of the excitation frequency, the displacement is periodic at the excitation frequency. It has been observed that a sinusoidal stimulus within the response region of a single nerve fiber results in firings of that fiber which ''may be synchronous with a particular phase angle of the stimulus.''[27] This synchronization to the phase of the stimulus is generally called phase-locking and is observed as more firings occurring in response to the maximum positive value of the input, and few occurring in response to the maximum negative value.[27,30] The firing rate, as a function of time, must therefore approximate an instantaneous voltage or potential, and not energy. The distinction is that phase is preserved in a potential function, while it is not in an energy function. If we consider several fibers in the same small response region, what we observe in their combined time-varying firing rate is approximately the original sine wave, with a DC bias. While there is no way of determining whether such is the case within the auditory process, a bandpass filter applied to the time-dependent neuron firing rate would easily recover the stimulating function. The high-pass component of the filter is needed to remove the DC bias of the firing rate, and the low-pass component is needed to smooth the time variations of the firing rate.

We can relate the bandpass filtered potential function to the STFT phase. Since the filtered potential function is essentially the source sine wave, we may complete it to an analytic signal by letting the imaginary part be the Hilbert transform of the potential function.[11] Since we have shown that $F_h(\omega_0,T)$ is the unique filtered analytic signal, the analytic completion of the potential function must be that analytic signal up to an additive constant phase. Therefore the instantaneous phase of the analytic completion of the potential function is essentially the STFT phase, up to an additive phase offset.

It has been noted that the ability in humans to distinguish frequencies is incredibly good, considering the breadth of frequencies to which a single point on the cochlea responds. In fact, the ear (and we use the term to refer to the entire auditory process) can distinguish the relative frequency difference of two short tones a few Hertz apart if the tones are presented to the ear with a short time interval between them. For instance, the just noticeable frequency difference of tones near 500 Hz for a sensation level of 40 dB and duration greater than 200 ms is about 1 Hz, and for a 10 ms tone the resolution is about 20 Hz.[31] At this frequency, the effective rectangular bandwidth of the cochlear response is about 100 Hz[30]

$$\mathrm{ERB}=24.7(4.37F+1), \tag{60}$$

where F is the center frequency in kHz. This means that humans have the ability to distinguish relative frequencies of these short duration tones with an accuracy approximately 100 times better than the effective bandwidth of the filter. This is comparable to the 10 dB bandwidth example depicted in Table I, where the frequency accuracy of the the cross-spectral based estimate was approximately 400 times better than the frequency quantization of the Fourier transform.

If we consider the problem of binaural source localization, spectral phase can provide accurate timing information for each channel, and therefore an accurate estimate of the time offset between the two channels. We present as evidence the impulse example of Fig. 2(d) and the collapsed speech excitation example of Fig. 8(d). In the impulse example, the entire STFT surface was re-mapped to a single point in time, which was precisely the time of the impulse. This means that the exact time of the impulse was recovered for each point (ω,T) on the STFT surface which responded to the stimulus. If we consider the excitation example of Fig. 8(d), we see that the glottal excitation was re-mapped to essentially an impulse. The question is with what precision can the time of excitation be estimated. If the precision of the estimate is ΔT, then a crude upper bound of the error in the time offset of the two channels is $2\Delta T$. For the example represented by Fig. 8(d), the precision in the estimated excitation time is about 150 ms, resulting in a crude maximum estimated channel timing error of 300 ms. We can compare these estimates with the just noticeable interaural time delay estimated at about 50 ms for frequencies below 1.5 kHz, as estimated from a just noticeable sensation of direction, which is about 5 deg for frontal incidence.[31] While the estimated error is not as good as human performance, it is also not based on an optimized process.

VI. CONCLUSIONS

We have presented a process, based on spectral phase, which effectively removes the blurring of the STFT surface by re-mapping the surface components to the FM functional representations of individual FM components of a multicom-

ponent signal. The process has been shown analytically to converge to the desired representation in the case of linear-FM signals, and an argument has been presented that the system must converge to the desired representation for signal components which may be represented as piecewise linear-FM. The results are valid, independent of the analysis window used, assuming a simple separability condition. The methods have been applied to speech to effectively isolate excitation and estimate formant frequencies instantaneously in both time and frequency. By modeling the speech signal at excitation and resonance, we have defined indicator functions, based on the mixed phase partial derivatives, which may be used to determine specific components on the STFT surface which are the response of excitation or resonance. By re-mapping the excitation indicated surface components, it was possible to conclude that the group delay of the vocal tract is essentially constant, making it possible to estimate the group delay of a transmission channel and phase equalize the signal to remove that channel group delay.

Note added in proof. It has recently come to the author's attention that reference should have been made to a method based on averaged group delay and instantaneous frequency mentioned by Kodera *et al*. as a variant of the methods proposed by them at least as early as 1986 (K. Kodera, R. Gendrin, and C. de Villedary, ''Analysis of time-varying signals with small BT values,'' in *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. ASSP-34, pp. 64-76, 1986.)

[1] J. Flanagan, *Speech Analysis Synthesis and Perception*, 2nd ed. (Springer-Verlag, New York, 1972).
[2] A. J. Gibbs, ''The Design of Digital Filters,'' in *Australian Telecommunication Research Journal*, Vol. 4, pp. 29–34, 1970, reprinted in *Digital Signal Processing*, edited by L. R. Rabiner and C. M. Rader (IEEE Press, New York, 1972), pp. 35–42.
[3] J. B. Thomas, *An Introduction to Statistical Communication Theory* (Wiley, New York, 1969).
[4] D. J. Nelson, ''Cross-Correlation and the Affine Invariant Pseudometric,'' Internal Technical Report, 1983.
[5] D. Nelson, ''Special Purpose Correlation Functions for Improved Signal Detection and Parameter Estimation,'' in *Proceedings of IEEE Conference on Acoustics, Speech and Signal Processing*, Minneapolis, MN, April, 1993, pp. 73–76.
[6] D. J. Nelson, ''Special Purpose Correlation Functions,'' Internal Technical Report, 1988.
[7] D. Friedman, ''Formulation of a Vector Distance Measure for the Instantaneous-Frequency Distribution (IFD) of Speech,'' in *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing*, Dallas, TX, 1987, pp. 1748–1751.
[8] D. Nelson and J. Pencak, ''Pitch based methods for speech detection and automatic frequency recovery,'' in *Proceedings of the SPIE Advanced Signal Processing Algorithms Conference*, San Diego, CA, 1995, pp. 92–100.
[9] T. Abe, T. Kobayshi, and S. Imai, ''Robust pitch estimation with harmonics enhancement in noisy environments based on instantaneous frequency,'' in *International Conference on Spoken Language Processes (ICSLP 96)*, Philadelphia, PA, 1996, pp. 1277–1280.
[10] D. J. Nelson and W. Wysocki, ''Cross-spectral methods with an application to speech processing,'' in *Proceedings of the SPIE Conference on Advanced Signal Processing Algorithms*, Denver, CO, July, 1999.
[11] L. Cohen, *Time-Frequency Analysis* (Prentice Hall, Englewood Cliffs, NJ, 1995).
[12] S. M. Kay, ''Statistically/computationally efficient frequency estimation,'' *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing*, New York, 1988, pp. 2292–2295.
[13] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1975).
[14] P. S. Murthy and B. Yegnanarayana, ''Robustness of Group-Delay-Based Method for Extraction of Significant Instants of Excitation from Speech Signals,'' IEEE Trans. Speech Audio Process. **7**, 609–619 (1999).
[15] P. S. Murthy and B. Yegnanarayana, ''Formant Extraction from Phase using Weighted Group Delay Function,'' Electron. Lett. **25**, 1609–1611 (1989).
[16] S. Umesh and D. J. Nelson, ''Computationally efficient estimation of sinusoidal frequency at low SNR,'' in *IEEE Conference on Acoustics, Speech and Signal Processing*, Atlanta, 1996, Vol. 5, pp. 2797–2800.
[17] S. H. Nawab and T. F. Quatieri, ''Short-Time Fourier Transform,'' in *Advanced Topics in Signal Processing*, edited by J. S. Lim and A. V. Oppenheim (Prentice-Hall, Englewood Cliffs, NJ, 1988).
[18] P. Maragos, J. F. Kaiser, and T. F. Quatieri, ''On amplitude and frequency demodulation using energy operators,'' IEEE Trans. Signal Process. **41**, 1532–1550 (1993).
[19] L. Cohen, ''Time-Frequency Distributions—A review,'' Proc. IEEE **77**, 941–981 (1989).
[20] R. Schmidt, ''A signal subspace approach to multiple emitter location an spectral estimation,'' Ph. D. dissertation. Dept. of Elect. Eng., Stanford University, Nov. 1981.
[21] R. Schmidt, ''New mathematical tools in direction finding and spectral analysis,'' Proc. SPIE, 7–19 (1983).
[22] B. S. Atal and S. L. Hanauer, ''Speech analysis and synthesis by linear prediction of the speech wave,'' J. Acoust. Soc. Am. **50**, 637–655 (1971).
[23] B. Yegnanarayana, ''Formant extraction from linear-prediction phase spectra,'' J. Acoust. Soc. Am. **63**, 1638–1640 (1978).
[24] L. Welling and H. Ney, ''Formant estimation for speech recognition,'' IEEE Trans. Speech Audio Process. **6**, 36–48 (1998).
[25] M. D. Plumpe, T. F. Quatieri, and D. A. Reynolds, ''Modeling of the glottal flow derivative waveform with application to speaker identification,'' IEEE Trans. Speech Audio Process. **7**, 569–586 (1999).
[26] Lord Rayleigh, ''On our perception of sound direction,'' Philos. Mag. **13**, 214–232 (1907).
[27] R. M. Warren, *Auditory Perception* (Prentice-Hall, Englewood Cliffs, NJ, 1995).
[28] G. von Bekésy, *Experiments in Hearing* (McGraw-Hill, New York, 1960).
[29] Helmholtz, *On Sensation of Time as a Physiological Basis for the Theory of Music* (2nd English Edition, rendered conformal to the last German Edition of 1877), translation by Alexander J. Ellis (Dover, New York, 1954).
[30] B. C. J. Moore, Ed., *Hearing*, 2nd ed. (Academic, New York, 1995).
[31] E. Zwicker and H. Fastl, *Psycho-acoustics*, 2nd ed. (Springer, New York, 1998).

Quantitative assessment of vocal development in the zebra finch using self-organizing neural networks

Petr Janata[a)]
Department of Organismal Biology and Anatomy, University of Chicago, Chicago, Illinois 60637

(Received 19 April 2001; revised 14 August 2001; accepted 20 August 2001)

To understand the mechanisms of song learning by songbirds it is necessary to have in hand tools for extracting, describing, and quantifying features of the developing vocalizations. The extremely large number of vocalizations produced by juvenile zebra finches and the variability in these vocalizations during the sensorimotor learning period preclude manual scoring methods. Here we describe an approach for classification of vocalizations produced during sensorimotor learning based on self-organizing neural networks. This approach allowed us to construct probability distributions of spectrotemporal features recorded on each day. By training the network with samples obtained across the course of vocal development in individual birds, we observed developmental trajectories of these features. The emergence of stereotypy in sequences of song elements was captured by computing the entropy in the matrices of first- and second-order transition probabilities. Self-organizing maps may assist in classifying large libraries of zebra finch vocalizations and shedding light on mechanisms of vocal development. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1412446]

PACS numbers: 43.80.Ka [WA]

I. INTRODUCTION

Juvenile zebra finch (*Taeniopygia guttata*) males undergo a period of vocal development between ~30 and 90 days of age during which the spectrotemporal properties of their vocalizations change significantly. The developmental progression is typically divided into three stages: subsong, plastic song, and crystallized song (Arnold, 1975; Zann, 1996). During subsong, which lasts from approximately 30 to 50 d, the vocalizations are generally quiet, sustained, and without regular repetition of identifiable spectral features. Plastic song (50–80 d) is characterized by the emergence of identifiable song elements, such as harmonic stacks, whistles, and frequency sweeps that are stable across many bouts of singing. Additionally, the ordering of song elements begins to assume a more stable structure. Finally, the adult crystallized song consists of a set of spectrotemporal song elements, referred to as notes and syllables, which are arranged into fixed sequences called motifs, phrases, or strophes.

The morphology of notes, syllables, and motifs is usually quantified by human observers along several feature dimensions (Scharff and Nottebohm, 1991), and song elements are classified based on these features. Such manual approaches are extremely labor intensive. Partially automated methods have been developed for classification of song elements (Anderson *et al.*, 1996; Kogan and Margoliash, 1998). Nonetheless, these approaches still require selection of templates and training based on pre-selected vocalization examples. Another method for automated feature extraction has been developed recently for quantifying the similarity of songs, e.g., songs produced by a tutor and a pupil (Tchernichovski *et al.*, 2001, 2000). The advantage of the latter method is that it derives similarity indices from a set of derived spectrotemporal features without making any assumptions about song-element boundaries. Automated methods have the advantage of eliminating human subjectivity from the similarity judgments as well as their ability to quickly compare multiple exemplars. Both human and automated classification methods work well in the case of crystallized song which is characterized by a relatively circumscribed set of distinct song elements.

The quantification and classification of subsong and plastic song vocalization features present significant challenges, however. The increased variability in spectrotemporal characteristics of subsong and plastic song, as well as the extremely large number of vocalizations produced on any given day, preclude human scoring. Consequently, selection of ''representative'' exemplars from these developmental stages for use in behavioral or neurophysiological experiments is somewhat idiosyncratic. An alternative approach would be to segment and classify, with minimal human intervention, the entire corpus of vocalizations recorded for an individual bird. Such an approach would allow one to determine the prevalence of a spectrotemporal pattern on any given day, and to track the emergence and disappearance of spectrotemporal patterns across vocal development. Selection of representative vocalization exemplars could then be based on objective statistical principles.

As an initial step toward this goal, we decided to establish the feasibility of using a self-organizing neural network algorithm to extract and cluster the spectrotemporal patterns encountered across the various stages of vocal development in individual zebra finches. Such an approach has been used to classify cries of human infants (Schonweiler *et al.*, 1996), speech sounds in general (Leinonen *et al.*, 1993, 1992), as well as musical instrument timbres (Toiviainen, 1996).

[a)]Present address: Department of Psychological and Brain Sciences, 6207 Moore Hall, Dartmouth College, Hanover, NH 03755. Electronic mail: petr.janata@dartmouth.edu

TABLE I. Summary of birds used in the experiment.

Bird	Age (d) at isolation	Age at last recording	Total #fragments analyzed
nj6	29	51	51 226
nj7	31	41	122 545
nj9	34	51	154 319
nj10[a]	30	91	562 077
nj11[a]	30	162	523 414
nj12	30	44	219 880
nj13	30	49	159 268
nj14	30	43	99 702
nj17[a]	34	82	567 558
nj19	30	50	222 925
nj22	37	53	199 198
nj23	54	58	60 168
nj24	55	62	143 728
nj25	60	65	49 016
nj26	34	46	176 378
zf_bk480	120+		41 200
zf_bk520	120+		15 056
zf_bk526	120+		48 191

[a]Indicates a juvenile whose vocalizations were recorded through adulthood.

II. METHODS

A. Experimental animals

The vocalizations of 18 male zebra finches (15 juveniles, 3 adults) were studied. Twelve juveniles were obtained from their home cages in our breeding colony at approx. 30 d of age (range: 30–37 d, mean 32 d). These birds were used for companion neurophysiological experiments. Each of these birds was removed at a different stage of vocal development and not returned to the experiment. Thus all recordings were obtained prior to any neurophysiological recording. The vocal development of three of these birds (nj10, nj11, nj17) was tracked into adulthood. Three additional juveniles were removed from their home cages between 54 and 60 days of age, several days prior to neurophysiological experimentation. As their vocalizations were recorded for several days, their data were included in order to increase the sample size in the 55–65 d age range. The vocalizations of three adult birds (>120 d) were recorded for 4–9 days so that the data analysis procedures described below could be assessed and tested using crystallized song. The age and duration of isolation of each bird is summarized in Table I.

Until the time of removal, juveniles were housed with both parents and any siblings. Following removal from the home cage, each bird was housed alone in a sound-attenuating chamber (Industrial Acoustics Corp.) with unrestricted access to food and water, and was maintained on a 14/10 h light/dark cycle. All animals were housed and treated according to protocols approved by the University of Chicago Institutional Animal Care and Use Committee.

B. Song collection

Sounds in the chamber were monitored continuously by means of a microphone (Model 33-2011, Realistic) suspended above the cage in the sound-isolation box. The signal was amplified and filtered (500 Hz high-pass; 10 000 Hz low-pass) with custom-built electronics (JFI Electronics, University of Chicago). The signals were digitized with 16-bit resolution at 20 000 samples/s (atMIO16x card, National Instruments) using custom software (Amish Dave, University of Chicago). Those signals that exceeded a specified amplitude threshold [typically twice the ambient root-mean-square (rms) of the signal] at least once during a 30 ms window in 10 out of 12 consecutive windows were recorded as an entry to computer disk. Recording of the entry stopped when the signal failed to cross threshold for 300 ms. Both short (<1 s) and long (>20 s) vocalization periods were captured with these settings. On rare occasions, the recording system would fail, resulting in gaps of one or two days in the vocalization database for any given bird.

Prior to the automated data analysis of the vocalizations, the spectrograms of all entries in the data files were visually inspected. Many entries consisted primarily of artifacts, e.g., cage noises, wing-flapping, and rustling of the food dish. These entries were excluded from the final dataset. Although this step was extremely time consuming, typically requiring 1–2 h of manual scoring for each day's vocalizations from a single bird, it was necessary in order to reduce the size of the original dataset to fit within computational constraints. Final reduced datasets ranged in size from several hundred megabytes to several gigabytes. The overall duration of identified song fragment sequences extracted from these datasets averaged 6.5 h/bird.

C. Automatic song parsing

All data analyses in this report were scripted in MATLAB (Mathworks, Natick, MA), and used functions in the Signal Processing, Neural Networks, and Statistics Toolboxes. Analyses were performed using a computer with 500 Mb RAM, running Linux on a 450 MHz Pentium II processor. The amplitude envelope of the waveform recorded in each entry was used to identify acoustic fragments that could serve as input data to the self-organizing map (SOM) algorithm. The signal was full-wave rectified and low-pass filtered (150 Hz) using a 5-pole Butterworth filter. A heuristic was empirically established to find those samples in the rectified and filtered waveforms that might constitute an amplitude peak (acoustic fragment). For most recorded entries, in which both sounds and extensive silent periods were present, the threshold criterion was set to be the median value in the signal. In some entries few silent periods were present, causing the criterion value to be set too high, resulting in the loss of many valid entries. Thus when the ratio of the mean and median values for the entry was <2, the criterion value was set to be one-fourth of the median value. Runs of samples that exceeded the threshold continuously for at least 10.5 ms were tagged as acoustic fragments that would enter into subsequent analyses. The continuity threshold was selected after inspecting the parsed data of several birds. It eliminated a large number of fragments that appeared unrelated to vocalizations while retaining the very short whistles that where observed in the vocalizations of some birds. In addition to retaining the waveform of each fragment, the onset and offset timing information about each fragment was preserved for subsequent use in identifying fragment sequences.

D. Self-organizing map training

A self-organizing neural network was presented with randomly selected exemplars from the bird's vocalization library in order to determine a mapping of acoustic features onto a vector of output units. The training set consisted of 20% of the total number of fragments for each bird taken across all days. To facilitate equal representation of vocalizations produced on different days, the maximum allowable number of fragments from each day contributing to the training set was equal across days. If the number of fragments for any given day was smaller than the daily allocation, all available fragments from that day were used. This happened only rarely, typically in the initial days of isolation in juveniles. Normally, a random sample of fragments was chosen from each day.

For two of the juvenile birds, whose vocalizations were monitored into adulthood (nj10, nj11), fragments recorded every second day between the ages of 50 and 60 d and every fifth day between 60 and 90 d were entered into the analyses. The sparser sampling was deemed adequate given the greater stability of vocalizations in these age ranges, and it precluded disproportionately weighting the random sample of training exemplars toward these ages.

1. Preprocessing of acoustical fragments

Input vectors to the neural network were time-frequency representations (spectrograms) of the acoustic fragments. Input vectors to the SOM were required to be of equal length, so it was necessary to specify a maximum fragment duration for each bird. Distributions of fragment duration (e.g., Fig. 2) showed that the proportion of long fragments was small. Thus in the interest of computational efficiency, fragments exceeding a criterion threshold were excluded from the analysis. For each bird, the threshold was fixed. Averaged across birds, the thresholds were 286.7 ± 58 ms (mean$\pm$std. dev.). On average, 98.78 ± 1.11% of the fragments for a bird were shorter than the criterion and included in the training and classification sets.

Each fragment in the training set constituted a single input vector to the training algorithm. First, the fragment was filtered with a fifth-order Butterworth filter (800 Hz high-pass; 8000 Hz low-pass settings). Fragments shorter than the established maximum fragment duration for each bird were padded with zeroes to achieve the proper length. Next, a spectrogram of the fragment was computed (specgram function in MATLAB) using a window length of 12.8 ms with 75% overlap between successive windows. A Hanning window was applied to each portion of the waveform before the Fast-Fourier Transform (FFT) was computed. In order to increase the temporal resolution in the input vector, while keeping the input vector's size tractable, values in successive pairs of frequency bins of the spectrogram were averaged, e.g., bin 1 and 2, bin 3 and 4, etc., thus yielding an effective frequency resolution in the spectrogram of 156.25 Hz/band. Only frequency bins in the range from 800 to 8000 were included in the input vector, as these were within the bandpass region of the filtering stage described above. The modified spectrogram was then ''unfolded'' to create a one-dimensional vector in which the spectra of successive time windows were laid end-to-end. Thus the length of the input vector corresponded to the number of averaged frequency bins (46 bins) multiplied by the number of time windows (e.g., 125 time windows), where the number of time windows differed for each bird depending on the fragment duration cutoff. Each input vector was normalized by the maximum value in that vector so that all input vector values would fall within a range from 0 to 1.

2. Network parameters

Self-organizing maps for each bird were created using the SOM functions in the Neural Networks Toolbox (Revision 1.3) in MATLAB. Briefly, the architecture consisted of a one-dimensional input layer connected to a one-dimensional output layer through a single layer of weights. Several output layer sizes and topologies were explored in several juvenile and adult birds to determine whether higher-dimensionality in the output layer facilitated classification of the song fragments. Output unit topology did not appear to influence the distributions of correlations between input vectors and the weight vectors connecting them to the winning output units. Therefore, for ease in displaying and interpreting the weight matrices, we settled on linearly arrayed output units. For adult birds we used 64 output units, and for juveniles we used 200. We used a smaller output vector for adults because the spectrotemporal variability in crystallized song is smaller, and presumably adequately represented with a smaller number of output units, than is the variability in juvenile subsong and plastic song.

Every input unit element was connected by a weight to every element in the output vector. Thus the weight matrix for a juvenile bird who had 5750 elements in the input vector contained 1 150 000 elements. Weights were initialized to random values. The weight matrix was updated through a competitive (''winner-take-all'') learning algorithm. Default values were used for the learning rates during the ''ordering'' phase (starting value of 0.9) and ''tuning'' phase (0.02). During the ordering phase, the size of the neighborhood in which weights were modified was gradually reduced in equal steps from the maximum distance between output units to a neighborhood of one unit. Similarly, the learning rate was reduced in equal steps from the starting value to the tuning phase value. Two-thirds of the training set were randomly selected and used in the ''ordering'' phase and the other third was used for the ''tuning'' phase.

E. Classification of song fragments and characterization of output unit loadings

Once the network had been trained for each bird, all identified fragments for the bird including those in the training set were classified. Each fragment was transformed into an input vector representation and correlated (Pearson correlation coefficient) with the weight vectors mapping the input vector to the output vector. The output unit associated with the vector of weights that correlated most highly with the fragment's spectrogram representation was chosen as the winning output.

Daily loading matrices were constructed for each bird by tallying the number of times each output unit was activated by that day's song fragments and dividing by the total number of fragments produced during that day.

F. Characterization of fragment sequences

1. Sequence identification

After the recordings had been parsed, sequences of acoustic fragments were identified as follows. Each entry in the recordings that contained multiple fragments was used for this purpose. A fragment was included in a sequence if its onset occurred within a criterion inter-fragment interval (IFI), measured from the offset of the preceding fragment. The criterion IFI was arbitrarily selected based on the IFI distribution for each bird, and was chosen to fall along the long tails of the distribution. In most cases, IFIs of 200 ms were used. This value was based on the empirical observation that the silence between fragments within motifs (in adults) or bouts (in juveniles) was less than 200 ms, and that longer intervals represented motif or bout boundaries.

Following SOM training and classification of every fragment, the identified fragment sequences were recoded as sequences of output units by replacing the identity of each fragment in the sequence with the output unit that it was classified under. In those cases where the fragment was not associated with an output unit of the network, i.e., if the duration of the fragment was too long, the fragment was assigned to an extra element in the output vector specifically used for these cases.

2. Transition probability matrices and entropy estimation

Once sequences of output unit activations had been identified it was possible to construct transition probability matrices (TPMs). For each day's vocalizations, a first-order TPM was constructed by tallying all first-order transitions. Each row in the TPM indexed the first of two sequence elements and the column indexed the second element. For example, the sequence {34, 10, 22, 34, 10} consists of four first-order transitions (34,10; 10,22, etc.) and would increment values in three elements of the TPM. The TPMs reflect the most frequent transitions between pairs of acoustic fragments. Second- and third-order TPMs were also computed for fragment triplets and quadruplets, respectively. Each row in a second-order TPM indexed a pair of fragments, and the different elements in that row indicated the overall likelihood of observing each of the different fragments following the particular pair of fragments. Third-order TPMs were similarly constructed.

Structure in each day's TPM, P, for each bird was quantified with the information theoretic measure of entropy, H:

$$H = \sum_{i=1}^{N} \sum_{j=1}^{N} -P_{i,j} \cdot \log_2 P_{i,j},$$

A normalized entropy value, H^*, was obtained using the total number of nonzero elements in the TPM:

$$H^* = \frac{H}{\log_2 g},$$

where g is the number of nonzero elements in the TPM.

The maximum possible normalized entropy value was $H^* = 1$, regardless of the number of nonzero elements in the TPM. This value would be obtained if all observed transitions were equally likely to occur. In order to estimate whether the structure in the observed TPMs differed from random probability distributions, we calculated simulated entropy values that would be expected given random sets of transition probabilities. For these simulations, we used 20 random vectors containing the same number of elements as the number of nonzero entries in the daily TPM for each bird.

III. RESULTS

A. Parser performance

The number of fragments identified for each bird is shown in Table I. Figure 1 shows examples of the parser's performance on the song of one juvenile (nj14) recorded at 32 and 43 days of age. Both the juvenile and the specific examples were selected randomly. The parser's performance did not always match the parsing by a human scorer. In cases where the amplitude threshold was set too high, a syllable consisting of two notes would be split into its constituent notes, whereas in other instances the syllable would be retained as a single unit. Similarly, if the heuristic resulted in a threshold that was set too low for a given entry, several seemingly separate elements would be grouped into a single element. Detailed inspection of the waveforms showed that often the transition between what appeared to be two closely apposed notes in a syllable were in fact separated by a very brief low-amplitude period. In such cases, the parser correctly identified the two acoustic events as separate, even though the expert human scorer of zebra finch song would tend to integrate the two events into a single, higher-order, event. Overall, we felt that the occasional ''errors'' of the parser were mitigated by its ability to efficiently process the immense volume of the data according to strict objective criteria. For example, the fragments shown in Fig. 1(B) represent merely 0.1% of the total number of fragments identified for this bird.

B. Duration statistics of zebra finch song fragments

One characteristic of crystallized zebra finch song is stable song element duration. This is exemplified in Fig. 2(A), in which the distributions of song fragment durations remained stable across four consecutive days. Not surprisingly, songs of different adults are characterized by different song fragment duration distributions [Figs. 2(A), (B)]. The cumulative duration distributions for 13 juveniles are shown in Fig. 2(C). Many of the distributions show peaks at various durations. However, the cumulative distributions mask any daily variation in fragment duration that may occur across the course of song development. Figure 2(D) illustrates the variability in fragment duration across song development in two juveniles. Both examples illustrate that by 75 days of

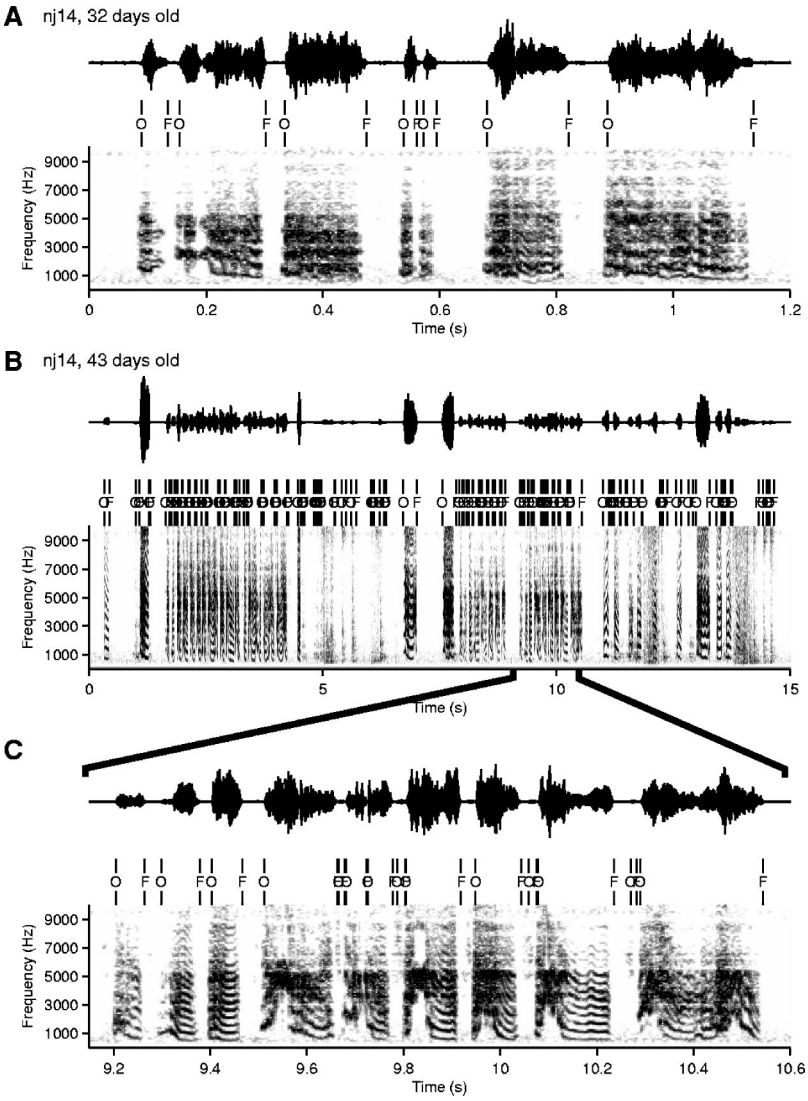


FIG. 1. Example of the parsing algorithm's performance. Two vocalization epochs from different developmental stages of the same zebra finch were selected at random. In each panel, the oscillogram at the top shows the amplitude fluctuations in the vocalization. Below it are shown the onset (O) and offset (F) marks for each fragment found by the parsing algorithm. The spectrogram is shown at the bottom. (A) A 1.2 s epoch of subsong recorded at 32 d of age. (B) A 15 s example of plastic song recorded from the same bird at 43 d of age. Cage noises, are evident in the recordings between 5 and 6 s (hopping), and again around 14 s (wing-flapping). The parsing algorithm had no way of distinguishing between vocalizations and cage noises. The 146 fragments shown in this epoch represent 0.1% of the total number of fragments identified for this bird. (C) An expanded view of the parser's output corresponding to a ~1.5 s segment in (B).

age the daily duration distributions contained a small number of distinct peaks. In the case of nj11, fragment durations were more uniformly distributed prior to day 45. The duration images also show that some peaks in the duration distributions shifted gradually along smooth trajectories after they initially formed, and some trajectories appeared to bifurcate.

C. Properties of the SOMs

The primary goal in utilizing a self-organizing network for the analysis of juvenile song fragments was to obtain an automatic classification of the various spectrotemporal characteristics present in the extremely large fragment dataset. Once the weight matrix linking the spectrogram representation with output categories had been established using a random subset (20%) of the fragments, the relative abundance (loading) of fragments in each output category was determined by correlating every fragment with each row in the weight matrix and assigning it to the output unit linked to the most highly correlated row in the weight matrix. The loading on each output unit could then be examined as a function of the bird's age.

The SOM approach was first tested on crystallized songs from adult zebra finches. Figure 3(A) illustrates a weight matrix for a single zebra finch, and the loadings on each output unit during each of the four days that this bird's song was recorded. Inspection of the weight matrix shows that clusters of adjacent output units coded similar features in the input vectors. For example, the rows of weights corresponding to output units 1–20 look very similar to each other, as do weights corresponding to output units 22–39, 40–42, and 57–64. Figure 3(B) illustrates that rows in the weight matrix form a very literal representation of the spectrotemporal features of the input vectors. Each panel shows a row (one-dimensional vector combining ''frequency'' and ''time'') of the weight matrix reshaped as a spectrogram (a frequency $\times$time matrix). The resulting spectrograms show identifiable zebra finch song elements. Since a row of weights maps the entire input vector onto a single output unit, the weights in a row of the matrix can be thought of as the components of a feature detector. Thus if an input vector, representing a set of spectrotemporal features, is correlated strongly with the row of weights, the output unit corresponding to that row of weights will be strongly activated.

Figure 4 shows weight matrices and daily weight matrix

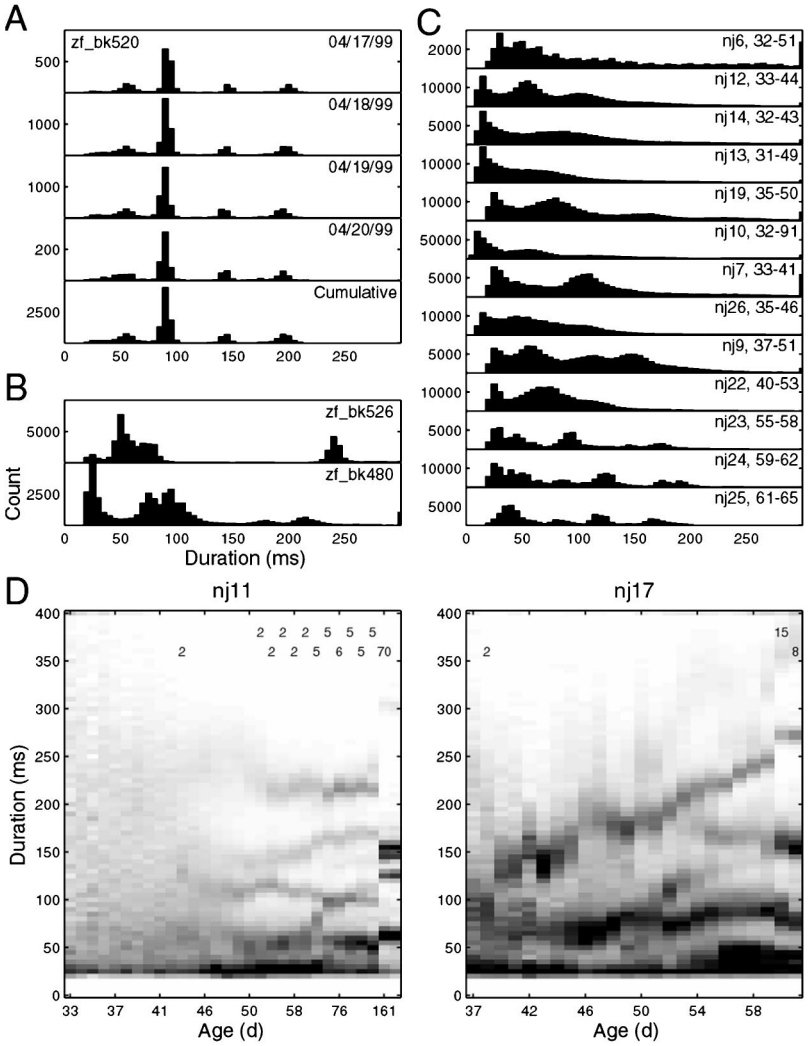


FIG. 2. Distributions of song fragment durations in adult and juvenile zebra finches. (A) Histograms showing the number of fragments observed for durations marked along the abscissa. Duration histograms for fragments recorded on each of four days from an adult zebra finch show little variation. (B) Cumulative duration histograms for two other adults. (C) Cumulative duration histrograms for 13 juvenile zebra finches. Each panel corresponds to the data for a single bird. The age range over during which the fragments contributing to the histogram were recorded is indicated at the top right of each panel. For example, 'nj9, 37–51', indicates fragments were recorded from bird, nj9, between 37 and 51 days of age. (D) Images of daily fragment duration distributions trace the evolution of fragment duration structure for two birds whose song was recorded across the period of sensorimotor learning. Gray scale intensity reflects the proportion of fragments for each duration on each day. The numbers at the top of each matrix indicate the number of days that were skipped between recording of fragments in the column with the number and the preceding column.

loadings for the three juveniles whose vocalizations were recorded over the course of their song development. Given the length of the input vector, the details of the spectrotemporal properties encoded by any given row of weights are lost when the weight matrix is viewed as a whole. Nonetheless, several properties of the weight matrix can be discerned at the coarse level. For instance, the lengths of the gray streaks in the weight matrices indicate which fragment durations different output units became sensitive to. Thus the gray-scale intensity profiles depict the overall temporal weighting functions applied to each fragment when calculating the best match. Short and long fragments tended to be represented at opposite ends of the output unit array.

Despite the coarse features represented in the overall view of the weight matrices, individual output units were sensitive to detailed spectrotemporal patterns (as shown in Fig. 3). The insets in Figs. 4(A) and (C) provide another example of the spectral features represented by the weights in four consecutive time windows. Figure 4(B) insets illustrate that nearby output units represent similar inputs, whose spectra during the same time windows differ primarily in the presence of a small peak at around 4 kHz in the plot of inset ii (see arrows). The daily output unit loading image for nj11 shows that the fragments activating the output units whose weight vector segments are plotted in the insets were recorded primarily between the ages of 47 and 51 d.

Together, the loading and weight matrices provide information about what the most common spectrotemporal features were in the song fragments produced at each stage of vocal development. Most striking about the loading matrices was the abruptness and magnitude with which some song features (as represented by the weight vectors) appeared. For example, for nj10, output units 170–190 were not loaded prior to 50 d, after which different members of the set were loaded for the remainder of the recording period. Conversely, output units that were loaded highly initially e.g., nj10, units 150–155 between 32 and 37 d, were loaded weakly or not at all following 60 d. For any given bird, the abrupt transitions did not all occur on the same day. For nj10, different spectrotemporal characteristics appeared at approximately 41, 44, 48, 50, and 54 days of age.

Figure 5 illustrates developmental trajectories captured by the weight matrices. For each bird, a portion of the weight matrix was selected that showed a gradual change in the distribution of loadings on adjacent output units across the course of several days. In other words, if adjacent output units represent subtle differences in their respective weight matrix rows (best-fitting spectrotemporal features), then loading of adjacent output units on successive days may represent gradual change in one or more features of song frag-

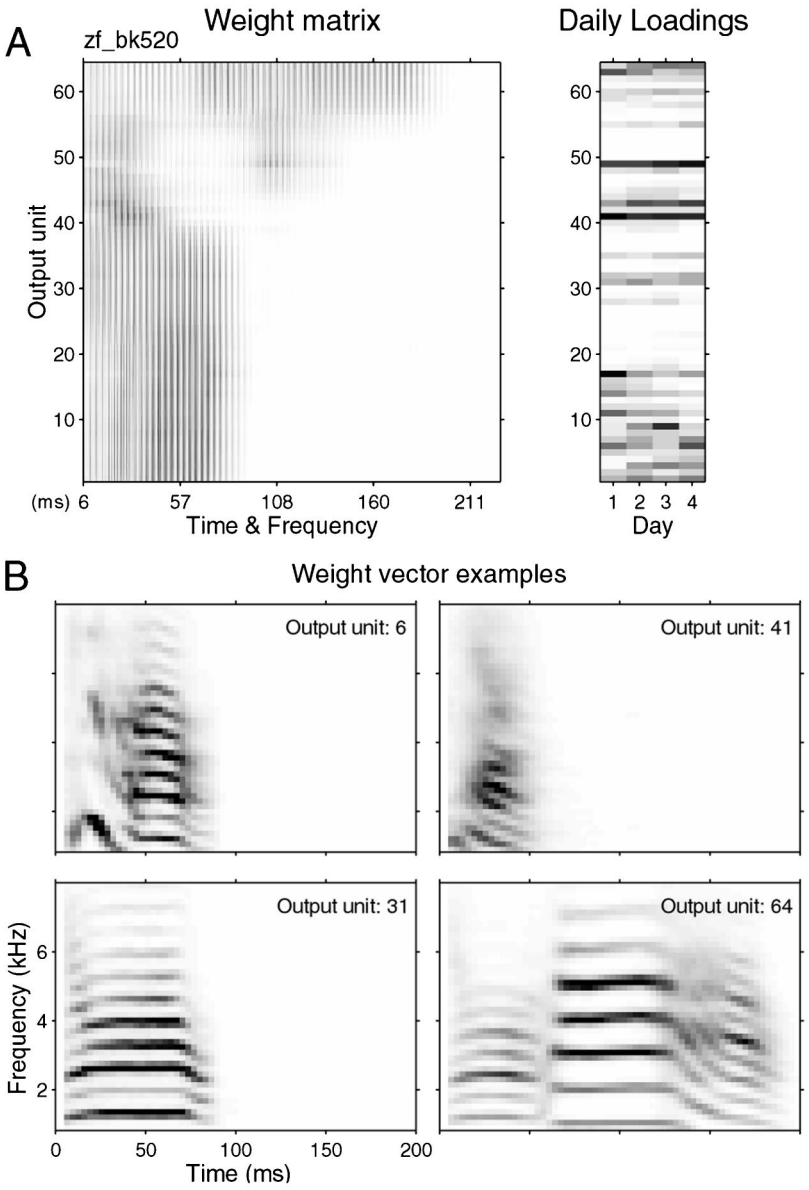


FIG. 3. SOM weight matrix and daily weight matrix output-unit loadings for an adult zebra finch. (A) The weight matrix shows the strength of association between all input units (columns labeled ''Time & Frequency'') and all output units (rows). Weight values are represented in gray scale intensity. The input layer corresponds to a spectrogram with the spectra of successive time windows laid end to end. The vertical striation in the weight matrices is a consequence of the ''unfolded spectrogram'' input representation. To the right of the weight matrix is a loading matrix which shows how often each output unit was activated on each day. Each row in the loading matrix corresponds to an output unit and is aligned with the corresponding output unit in the weight matrix to the left. The columns represent days. The gray scale intensity of each element in the matrix represents the proportion of fragments recorded on a particular day that was most strongly correlated with the output unit represented by that row. Darker values indicate a higher proportion. (B) Each panel shows the weight values of a row in the weight matrix in (A), rearranged into a two-dimensional spectrogram representation. These images make evident that the weights in the SOM adapt to represent specific features in the input vectors and that the output units serve as ''feature detectors'' for these specific features.

ments that most strongly activate those output units. Each of the song fragments shown in Fig. 5 was the fragment that correlated most highly with the particular output unit on the specified day. In the case of nj11, the harmonic stacks become more distinct with increasing age. The song fragments for nj17 show a more complex pattern of change, including both a lengthening of the component note elements and continued differentiation of the spectral features in the second half of the song fragment.

D. Analysis of produced sequences

Development of zebra finch song is characterized not only by the emergence and crystallization of the spectrotemporal features of individual song elements, but also by arrangement of these song elements into fixed, stereotyped sequences. In this study, song fragments were identified as belonging to the same sequence if the time between the end of one fragment and the start of the next was less than 200 ms. Classification of fragments into categories using the SOM facilitated the quantification of emerging structure in fragment sequences because of the sheer number of fragments and sequences that could be labeled automatically. All the sequences recorded on any given day were used to construct first-order TPMs which summarized for each output unit the likelihood that it would be followed by itself or some other output unit. Because many output units were not loaded on any given day, and because transitions were not observed between all possible pairs of output units, the TPMs were rather sparse for any given day. The most dense TPMs showed at least one transition for 12 500 out of 40 000 possible transitions, i.e., ~31% of the entries in a TPM had nonzero values. The number of overall fragments produced during a day and the number of nonzero entries in the TPM were significantly correlated (Fig. 6). A test of the difference in regression coefficients for the young and older birds showed that for any given number of fragments produced, however, the TPMs for older birds showed significantly fewer transitions (nonzero entries) $[F(1,240)=9.764,\ p$

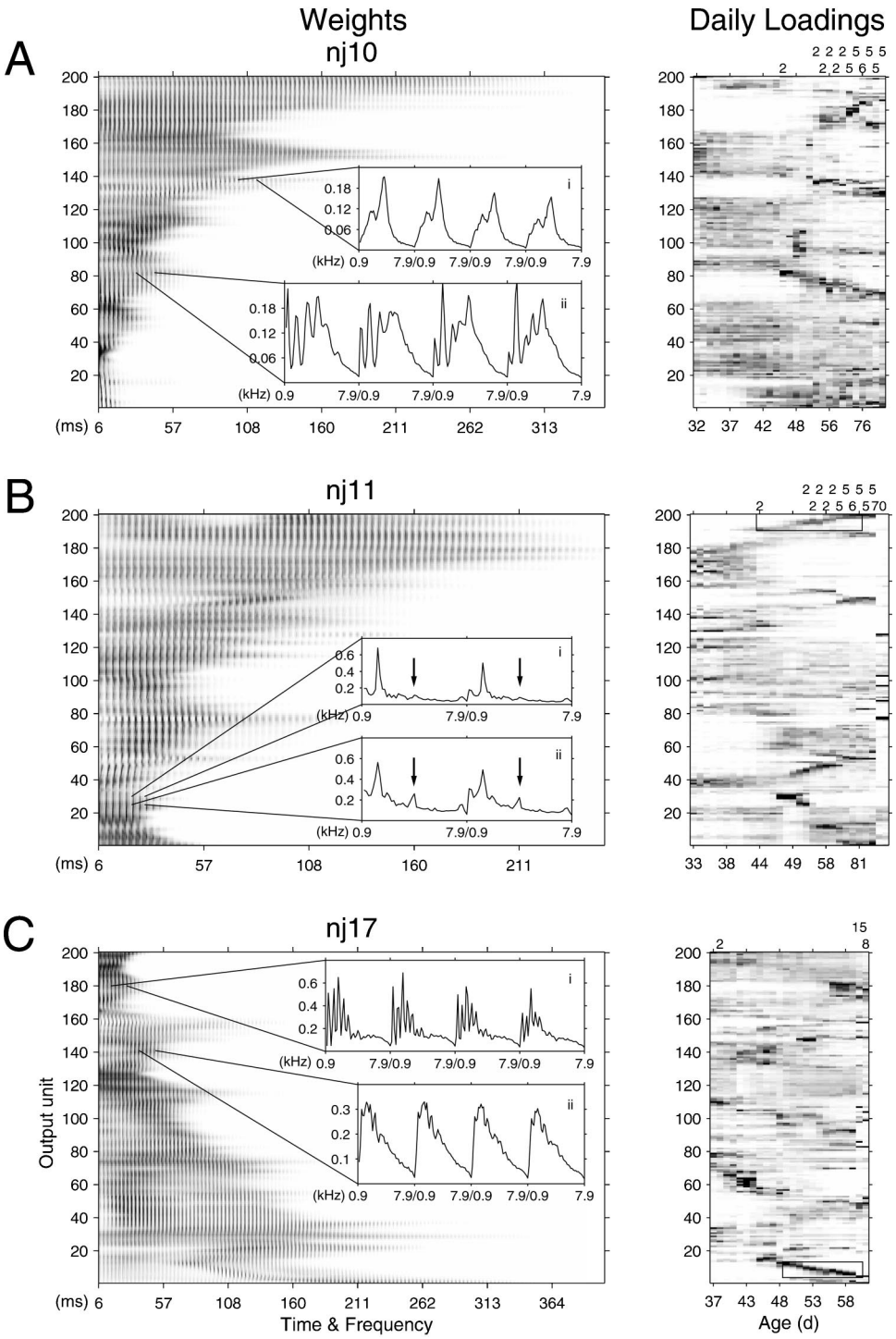


FIG. 4. SOM weight matrices and daily weight matrix output-unit loadings for three zebra finch juveniles whose vocal production was recorded across the bulk of the sensorimotor learning period. See Fig. 3 for a description of how to interpret the weight matrices and loading matrices. Each inset in the weight matrices plots the weights linking a section of the temporo-spectral input representation with an output unit. For example, inset A-i shows the spectral features in four successive time windows occurring at ~108 ms from fragment onset to which the subset of weights is sensitive. The matrices of daily output unit loadings (shown at the right) indicate that some output units represent song fragments produced in early stages of song development, whereas others encode fragments produced at later stages. For example, the output unit linked to the weights shown in inset C-i was primarily activated after the age of 56, whereas the output unit linked to the weights shown in C-ii was activated primarily by fragments produced between ages 43 and 48. The numbers above each loading matrix indicate the number of days that were skipped between recording of fragments in the column with the number and the preceding column. The boxes in the loading matrices for nj11 and nj17 enclose regions that form the basis for Fig. 5.

<0.002]. This indicated that a smaller set of transitions occurred more often in older birds, as would be expected of more stereotyped sequences.

We quantified the amount of structure in a TPM by computing the entropy in the TPM. Entropy is maximal if all transitions are equally likely to occur. Because the calculated entropy value was normalized with the number of nonzero elements in the TPM, i.e., the maximum entropy given the number of nonzero elements, the entropy values ranged from 0 to 1. Entropy in the daily loading matrices, i.e., in the distribution of probabilities of activating any given output unit increased until day 38 and then stabilized, indicating an increase in the diversity of activated output units. Entropy in the first- and second-order TPMs decreased with increasing age (Fig. 7, circles and diamonds). The entropy of the TPMs was compared with entropy values that would be obtained by assigning random probability values to the nonzero entries in the observed first-order TPMs (with the constraint that the probabilities sum to 1). Somewhat surprisingly, the entropy of the observed first-order TPMs was higher (0.9864 $\pm$0.0038 std. dev.) than for random first-order TPMs (0.9702 $\pm$0.0022 std. dev.) in the age range of 31–35 days. After 46 d of age, the entropy of the observed first-order TPMs was substantially lower than the entropy for random TPMs, and

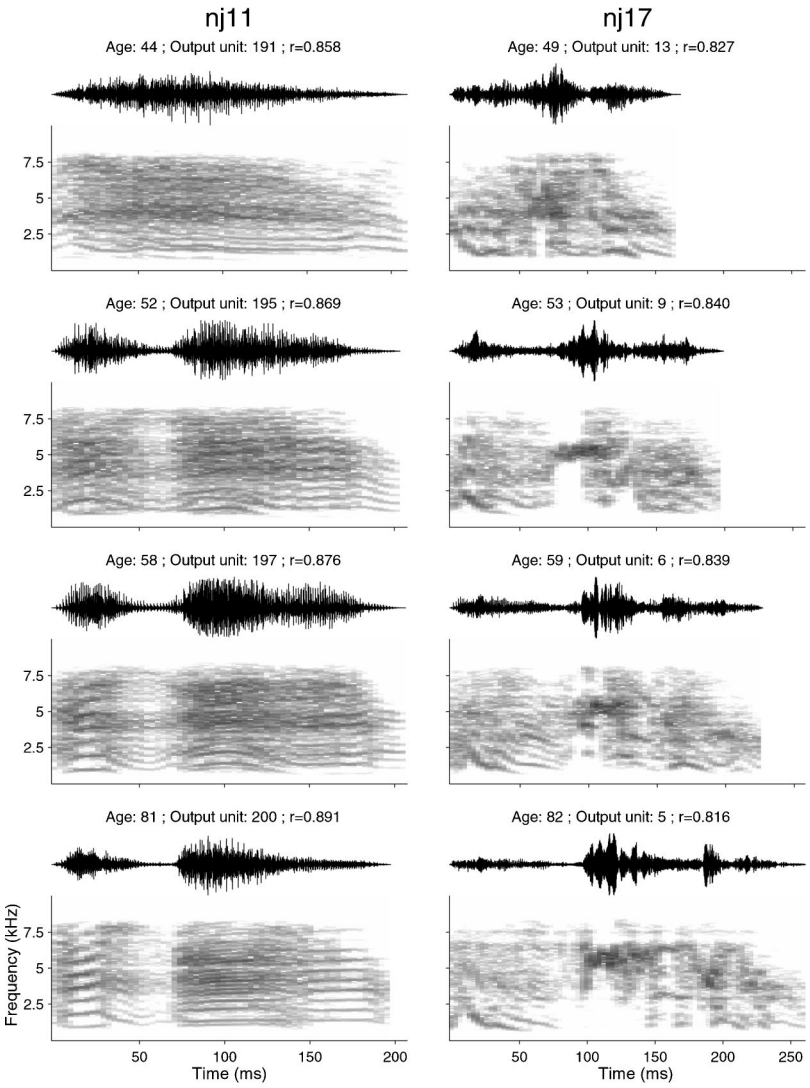


FIG. 5. Examples of song fragments that loaded nearby output units on successive days. Oscillograms and spectrograms on the left show the song fragment that was most highly correlated with the weights associated with the specified output unit on each of four days for bird, nj11. The age of the bird, output unit (row in the weight matrices shown in Fig. 4), and the magnitude of the correlation are specified in the title of each plot. Song fragments for bird, nj17, are shown on the right. Note the similarity and development of note features as the birds age (top to bottom).

decreased to ~0.7 in nj11 who was recorded until 162 days of age (data not shown). Second-order entropy also started to decrease after 46 days of age (Fig. 7, diamonds). The average TPM entropy for adults was 0.881 ± 0.003 (std. dev.), 0.948 ± 0.018, and 0.983 ± 0.009 for first-, second-, and third-order transitions, respectively.

IV. CONCLUSIONS

Using a simple input representation of zebra finch vocalizations (the amplitude component of FFT-based spectrograms), and a simple self-organizing neural network architecture consisting of a single weight layer and one-dimensional output vector, we generated maps of individual zebra finch vocalization histories. When reconstituted as spectrograms, rows of connection weights mapping the spectrotemporal input vector to output units appeared as plausible song elements. This indicated that the SOMs had extracted the most prominent spectrotemporal features in the song fragment database for each bird. Using the SOMs, automated classification of tens to hundreds of thousands of song fragments from individual birds enabled us to generate a statistical description of which features were present when during vocal development. The method identified the emergence and disappearance of spectrotemporal features that had come to be represented in the weight matrix. In many cases, adjacent output units of the SOMs were heavily loaded on successive days, forming identifiable trajectories in the daily output loading matrices. Trajectories in these matrices appear to represent development trajectories of spectrotemporal features of song elements.

Tens of thousands of sequences were automatically labeled and transition probabilities between sequence elements were calculated. The entropy in the first-order transition probability matrices decreased with increasing age of the bird, indicating that the ordering of song elements, as represented in the SOM output layer, became less random as the bird's vocalizations developed. This is in agreement with qualitative observations of increased sequence stereotypy as the zebra finch crystallizes his song. The drop in first-order TPM entropy, averaged across animals, around day 45 corresponds well to the observed juncture between subsong and plastic song stages of song development (Zann, 1996).

While the SOM approach provides a convenient means of reducing extremely large datasets of vocalizations, what insights into the vocal development process do the resulting SOMs and derived sequence entropy measures provide? As

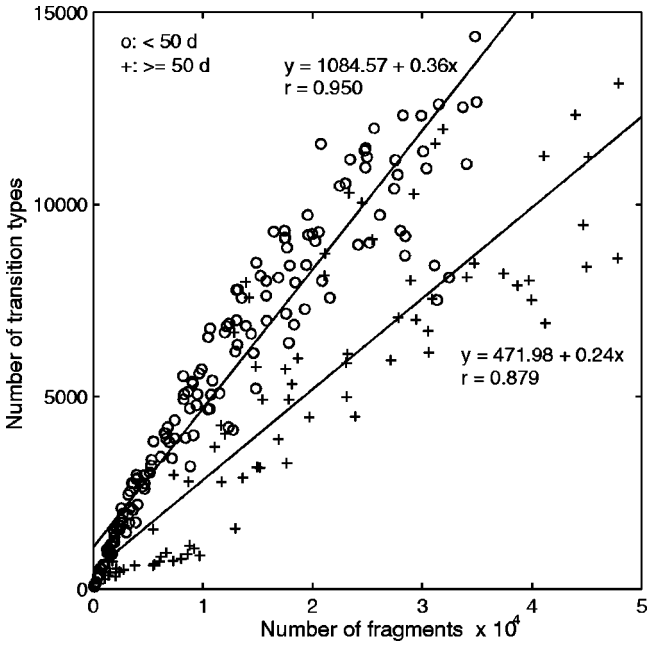


FIG. 6. Relationship between the number of fragments recorded during a day and the number of different first-order transitions between output units activated by those fragments. Circles represent data for fragments produced before 50 d of age and crosses correspond to fragments produced after 50 d of age.

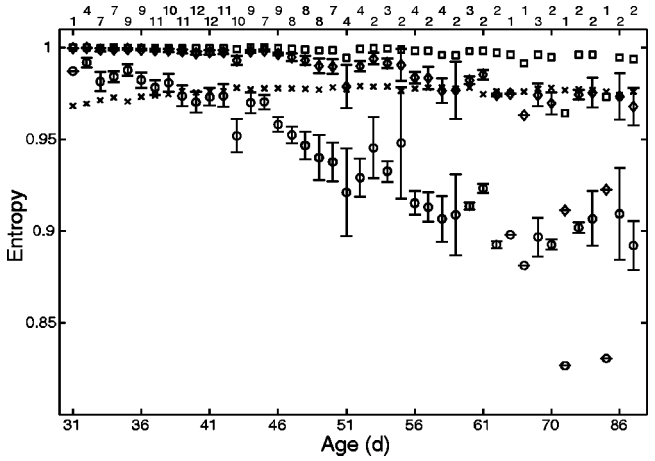


FIG. 7. Entropy of song fragment sequences as a function of age. Circles indicate average entropies of the daily first-order transition probability matrices (TPMs), diamonds correspond to entropy in second-order TPMs, and squares correspond to entropy in third-order TPMs. The entropy of a TPM was normalized using the number of nonzero entries in the TPM. If all observed transitions were equally likely, entropy would equal 1. Crosses indicate simulated entropy values that are obtained if the distributions of probabilities in the first-order TPMs are randomly determined using the number of non-zero entries in the observed data. Error bars represent ± 1 S.E.M. In order to reduce clutter in the figure error bars are not shown for the simulated and third-order data because they rarely extend beyond the bounds of the plotted symbols. The number of birds contributing to the entropy estimate for each day is shown above the plot.

indicated above, fragment sequences labeled using the SOM showed increased stereotypy with increasing age, mirroring qualitative descriptions of zebra finch vocal development. To our knowledge, the SOM results provide the first quantitative estimates of when changes in sequence structure occur, based on a nearly exhaustive sampling of the vocalization history. Aside from the ability to quantify changes in sequence structure, the SOM weight matrices and associated daily loading matrices suggest that two types of developmental phenomena are captured by the SOM approach. The first type represents developmental trajectories that arise from gradual changes in spectrotemporal features across several days. This type of trajectory is captured by virtue of the SOM algorithm modifying not only the weights between the input vector and the most highly activated output unit, but also the weights of neighboring units. This leads to a clustering of output units whose connection weights are very similar and represent similar spectrotemporal features in the input layer. The observation that adjacent output units were maximally loaded on successive days in the daily loading matrices indicates that spectrotemporal features were changing subtly across time.

The other type of phenomenon is the sudden appearance of new spectrotemporal features. This was observed as heavy output unit loading starting on one day and then continuing on successive days, with no loading of the output unit on previous days. In other words, the spectrotemporal features represented by the output unit came into sudden existence in the vocalization database. It was possible to represent such features in the weight matrices because fragments sampled equally from all stages of development were presented randomly to the neural network during the training phase. Thus all frequently occurring spectrotemporal feature categories had a chance of establishing themselves in the weight matrix.

The notion that the both the gradual trajectories and sudden loadings in the daily weight matrix loading maps represent actual modes of vocal development in the zebra finch derives from the recent work of Tchernichovski and colleagues (2001). In their analysis, zebra finch vocalizations are decomposed into four feature parameters: Wiener entropy, spectral continuity, pitch, and frequency modulation (Tchernichovski *et al.*, 2000). At any given developmental stage, a vocalization is described by the relative magnitudes of these four parameters. A similarity index for vocalizations recorded at different developmental stages is computed by comparing the distributions of values on these four parameters. When assessing the imitation of tutor song elements presented to juvenile birds under highly constrained operant learning conditions, they found that song learning, i.e., modification of the four song features, proceeded along ''direct'' and ''indirect'' routes (Tchernichovski *et al.*, 2001). In the ''direct'' imitation trajectories, the features changed gradually, whereas in the ''indirect'' trajectories, the pitch feature would change gradually until a critical point at which the period would suddenly double. Further work is needed to establish whether the trajectories observed by Tchernichovski *et al.* are similar to the trajectories appearing in the daily weight matrix loadings.

A potentially promising approach would be to merge elements of both methods. For example, input vectors could be built for each fragment from values on each of the four feature dimensions described above. The role of the SOM would be to extract the organization of this four-dimensional feature space across the developmental history of the individual bird. One advantage of this approach might be a re-

duction in the computation time resulting from a reduction in the size of the input layer because each time window in the fragment would be represented by only four feature parameters rather than a larger number of frequency bins.

In performing the analyses described in this paper we identified several steps that call for further improvement. First among them is eliminating the need to manually separate those recorded entries containing song from those containing predominantly calls and cage noises. In some cases, entries with cage noises and/or calls represented 50% of the recordings. This resulted in several problems. First, if entries containing cage noises and calls were parsed and added to the fragment database, the database became unmanageably large requiring several gigabytes of disk storage for each bird and several hundred megabytes of RAM for handling the data structures. More significantly, if allowed into the final dataset, cage noises and calls would comprise a disproportionate amount of the SOM training set, since training was based on a random sample of 20% of the acoustical fragments in the final dataset. Thus we found it necessary to discard entries containing primarily cage noises and calls. For a trained scorer this process took approximately 1 h for each day's recordings from a single bird.

In our analyses, we adopted an architecture with output units arranged along a single dimension. We did this primarily because the weight matrix is easier to look at and understand with this type of architecture. However, it is reasonable to assume that the optimal organization of spectrotemporal features in developing zebra finch song might be captured more accurately with higher dimensionality in the output layer. As a preliminary test of this idea, we used the data from an adult bird to train SOMs with either a one-dimensional output vector with 64 units, or a six-dimensional output vector with 2 units along each dimension (64 units total). Following training, we correlated each fragment with the respective weight matrices and constructed the distributions of maximal correlations. There appeared to be no difference in the correlation distributions (data not shown) so we did not pursue this issue further and chose instead to illustrate the method on the one-dimensional case.

Given our initial results, we believe that self-organizing neural networks promise to be a useful tool for the objective categorization of zebra finch vocalizations recorded over the course of vocal development.

ACKNOWLEDGMENTS

The research was supported by the following NIH grants: F32 NS10395 to P.J., R01 NS25677 to Daniel Margoliash, and P50 NS17778-18 to Jamshed Bharucha. The help of Alphi P. Elackattu in the screening of recordings was greatly appreciated. Patrice Adret and Timothy Q. Gentner provided helpful comments on an earlier version of the manuscript.

Anderson, S. E., Dave, A. S., and Margoliash, D. (**1996**). ''Template-based automatic recognition of birdsong syllables from continuous recordings,'' J. Acoust. Soc. Am. **100**, 1209–1219.

Arnold, A. P. (**1975**). ''The effects of castration on song development in zebra finches (*Poephila guttata*),'' J. Exp. Zool. **191**, 261–278.

Kogan, J. A., and Margoliash, D. (**1998**). ''Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study,'' J. Acoust. Soc. Am. **103**, 2185–2196.

Leinonen, L., Hiltunen, T., Torkkola, K., and Kangas, J. (**1993**). ''Self-organized acoustic feature map in detection of fricative-vowel coarticulation,'' J. Acoust. Soc. Am. **93**, 3468–3474.

Leinonen, L., Kangas, J., Torkkola, K., and Juvas, A. (**1992**). ''Dysphonia detected by pattern recognition of spectral composition,'' J. Speech Hear. Res. **35**, 287–295.

Scharff, C., and Nottebohm, F. (**1991**). ''A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: Implications for vocal learning,'' J. Neurosci. **11**, 2896–2913.

Schonweiler, R., Kaese, S., Moller, S., Rinscheid, A., and Ptok, M. (**1996**). ''Neuronal networks and self-organizing maps: New computer techniques in the acoustic evaluation of the infant cry,'' Int. J. Pediat. Otorhinolaryngol. **38**, 1–11.

Tchernichovski, O., Mitra, P. P., Lints, T., and Nottebohm, F. (**2001**). ''Dynamics of the vocal imitation process: How a zebra finch learns its song,'' Science **291**, 2564–2569.

Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B., and Mitra, P. P. (**2000**). ''A procedure for an automated measurement of song similarity,'' Anim. Behav. **59**, 1167–1176.

Toiviainen, P. (**1996**). ''Optimizing auditory images and distance metrics for self-organizing timbre maps,'' Journal of New Music Research **25**, 1–30.

Zann (**1996**). *The Zebra Finch* (Oxford University Press, New York).

PROGRAM OF

The 142nd Meeting of the Acoustical Society of America

Greater Ft. Lauderdale–Broward County Convention Center • Fort Lauderdale, Florida
3–7 December 2001

NOTE: All Journal articles and Letters to the Editor are peer reviewed before publication. Program abstracts, however, are not reviewed before publication, since we are prohibited by time and schedule.

MONDAY MORNING, 3 DECEMBER 2001 ROOM 223, 9:25 A.M. TO 12:00 NOON

Session 1aBB

Biomedical Ultrasound/Bioresponse to Vibration: High Intensity Focused Ultrasound and Bubbles

Michael R. Bailey, Chair
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

Chair's Introduction—9:25

Contributed Papers

9:30

1aBB1. Modeling and direct visualization of temperature rise induced by high-intensity ultrasound in tissue. Vera Khokhlova (Dept. of Acoust., Phys. Faculty, M. V. Lomonosov Moscow State Univ., Moscow 119899, Russia), Nathan Miller, Ryan Ollos, Roy Martin, Michael Bailey (Appl. Phys. Lab., Univ. of Washington, Seattle, WA), Yahya Mohammadian, and M. Naghavi (Texas Heart Inst. and the Univ. of Texas Med. School, Houston, TX)

High-intensity focused ultrasound (HIFU) creates localized heating deep in tissues which can be used to necrose tumors or cauterize vessels. Imaging of the temperature field in tissue is important for guiding HIFU treatment. Temperature rise in excised and degassed bovine liver exposed to high-intensity focused ultrasound was visualized experimentally and simulated numerically. An infrared camera, which records surface temperature only, was used to measure spatial temperature distribution. Two blocks of tissue were stacked, and their interface was placed along the axis of the ultrasound beam. A single element concave transducer (2 MHz, 35-mm aperture, 51-mm radius of curvature) was used. After exposure to ultrasound, the upper piece was immediately removed, and the temperature on the axial plane was infrared imaged. The absorption coefficient of liver was measured and then used for numerical simulations. The theoretical model employs a KZK-type equation for the acoustic pressure field combined with a bioheat equation. It is shown that experiment and theory agree well on the location, shape, and dimensions of the heated region. The dependence of absorption coefficient in liver on exposure to ultrasound, to air, and to degassing process was studied. [Work supported by NIH Fogarty, NSF, and CRDF.]

9:45

1aBB2. An innovative synthetic tissue-mimicking material for high-intensity focused ultrasound. Cyril Lafon (Ctr. for Industrial and Medical Ultrasound, Seattle, WA 98105), Oleg A. Sapozhnikov (Moscow State Univ., Moscow, Russia), Peter J. Kaczkowski, Shahram Vaezy, Misty Noble, and Lawrence A. Crum (Ctr. for Industrial and Medical Ultrasound, Seattle, WA)

A dosimetry study of the high-temperature and pressure regimes involved in high-intensity focused ultrasound (HIFU) requires experiments on biological tissues because no synthetic tissue-mimicking phantom is available. Unfortunately, the development of coagulative lesions cannot be observed in real-time in opaque tissues. Furthermore, the natural heterogeneous structure of tissue complicates direct comparison with numerical models. In this study, a new optically transparent phantom is evaluated. It is principally composed of a polyacrylamide gel, and includes a thermally sensitive indicator protein that becomes optically diffusive when denatured. Various tests were undertaken to characterize the acoustical, thermal, and optical properties of this material for a range of protein concentrations. The attenuation coefficient can be usefully modified by adjusting the quantity of embedded proteins to permit some selection of acoustic regime. It is also possible to emphasize cavitation activity at lower BSA concentrations, or thermal effects at higher concentrations. This new phantom adequately matches tissue for most of the measured parameters and facilitates the study of the HIFU bioeffects. [Work supported by ONR Grant No. N00014-96-1-0630 and NIH Grant No. HL64208-02.]

10:00

1aBB3. A multi-channel high-intensity focused ultrasound (HIFU) system for image-guided therapy (IGT). Peter J. Kaczkowski, Shahram Vaezy, Roy Martin, Lawrence A. Crum (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, peter@apl.washington.edu), and George Keilman (Sonic Concepts, Woodinville, WA 98072-7029)

Recent success in using high-intensity focused ultrasound (HIFU) for surgical applications has stimulated the development of image-guided systems for noninvasive therapy. With the goal of developing an all-ultrasound modality, we have developed a clinical HIFU system comprising an integrated transducer probe that permits precisely registered HIFU therapy delivery and high-quality ultrasound imaging, a multi-channel signal generator and rf amplifier system, and a software program that provides the clinician with a graphical overlay of the ultrasound image and therapeutic protocol controls. Electronic phasing of the 2-MHz HIFU annular array allows adjusting the focus within the range of about 4–12 cm from the face. A central opening in the transducer permits mounting of a commercial medical imaging scanhead (ATL P7-4) that is held in place within a special housing. This mechanical fixture ensures precise registration between the HIFU transducer axis and the image plane of the imaging probe. We will present a description of the various system components along with experimental *in vitro* results of therapy targeting and lesion visualization.

10:15

1aBB4. An optimal protocol for shock wave delivery by the dual-pulse lithotripter. Dahlia L. Sokolov, Michael R. Bailey, and Lawrence A. Crum (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., 1013 NE 40th St., Seattle, WA 98105, dsokolov@apl.washington.edu)

Dual pulses, generated by opposing, confocal, and simultaneously triggered electrohydraulic shock wave sources, can be used to localize and intensify cavitation for application to lithotripsy. It has been reported that the dual-pulse lithotripter (DPL) may be a safer and more effective lithotripter [Sokolov *et al.*, J. Acoust. Soc. Am. **100** (2001)]. Fragmentation of cement stones and hemolysis in dilute suspensions of human red blood cells were assessed at several positions along the focal axis of the DPL. Pulse repetition frequency was varied from 0.5 to 2 Hz and charging voltage was varied from 12 to 20 kV. At 1 Hz and 15 kV, the number of stone fragments $>$ 1.5 mm, at the focus, remained the same while hemolysis, 2 cm away, decreased by one-half, when compared to the conventional lithotripter at the same PRF and a higher voltage of 18 kV. By varying PRF and voltage, an optimal *in vitro* protocol for shock wave delivery that maximizes stone fragmentation at the focus and minimizes hemolysis away from the focus might be determined. Results support the hypothesis that cavitation plays a more significant role than shock waves in both stone fragmentation and cell damage in the dual-pulse lithotripter. [Work supported by NIH DK43881.]

10:30

1aBB5. Acoustic and cavitation field of an orthopedic shock wave device HMT Equitron. Parag Chitnis and Robin Cleveland (Aerosp. and Mech. Eng. Dept., Boston Univ., 110 Cummington St., Boston, MA 02215)

Extracorporeal shock wave therapy (ESWT) is a recent development in the treatment for orthopedic ailments. The SW devices are typically reduced versions of lithotripters. Measurements were taken on a HMT Equitron with a 35-mm focal length shock head. Acoustic measurements were taken with a PVDF membrane hydrophone. The average peak positive and peak negative pressures were observed to be 20 MPa and 12 MPa, respectively, with little dependence on the energy level. The focal spot was found to be 15 mm long and 16 mm wide. The cavitation field induced by SWs was measured using a passive cavitation detector (PCD) consisting of a 1-MHz focused transducer. Cavitation bubbles generated two acoustic emissions similar to the response observed in lithotripsy. The first emission coincided with the arrival of the shock waves and the second with the inertial collapse of the cavitation bubbles. The characteristic time (tc) was obtained by the time difference of the two signals. The region of strong cavitation field corresponded to the focal region of the acoustic field. The Gilmore equation was used to predict bubble response from the measured acoustic waveforms and the results were in agreement with PCD measurements. [Work supported by NIH.]

10:45–11:00 Break

11:00

1aBB6. Comparison of ultrasound scattering behavior of optison and a liquid perflourocarbon nanoparticle contrast agent. Michael Hughes, Jon Marsh, Ralph Furhop, Lori Chinen, Sam Wickline, and Greg Lanza (Washington Univ., St. Louis, MO 63110)

Useful ultrasound contrast enhancement has been obtained using liquid emulsion contrast agents developed in our laboratory [Evaluation of liquid perflourocarbon nanoparticles as a blood pool contrast agent utilizing power Doppler harmonic imaging, IEEE Ultrason. Symp. **00CH37121**, 1931–1934 (2000)]. While the physical basis for enhancement is unknown, it is reproducible, *in vivo* and *in vitro*. It is hypothesized that microbubble formation is not a significant factor in the acoustic behavior of these contrast agents; attenuation and backscatter (measured in a water bath at 37 °C) are uneffected by changes in hydrostatic pressure (from -50 to 200 mm Hg), acoustic pressure (0.038–3.0 MPa), and show no evidence of scattering agent destruction with ultrasound exposure time (2–80 s). The microbubble agent attenuation coefficient exhibits a single peak between 1 and 2 MHz. The exact peak location, height and width vary significantly with experimental conditions. In contrast, the emulsion does not exhibit any of these effects, which makes it appear very unlikely that microbubble formation through perflourocarbon phase conversion contributes in a measurable manner to the acoustic properties of the liquid perflourocarbon nanoparticle contrast agent over the range of experimental parameters in this study. [Work supported in part by HL-59865, N01-C0-07013-32.]

11:15

1aBB7. Dynamics and fragmentation of encapsulated gas bubbles. John S. Allen, III, Donovan J. May, and Katherine W. Ferrara (Dept. of Biomed. Eng., Univ. of California, Davis, Davis, CA 95616)

The development of ultrasound contrast agents has fostered an interest in the theoretical dynamics of encapsulated gas bubbles. Originally, micron-sized agents were constructed for diagnostic imaging purposes with a thin shell ($<$15 nm) of lipid, protein, or polymeric materials. New therapeutic applications for localized drug delivery have led to an interest and development of micron-size bubbles with a thick fluid shell ($\sim$500 nm). Previously some experimental acoustic levitation studies have investigated the capillary wave development on millimeter-size liquid shells. The nonlinear oscillations and fragmentation have not been extensively investigated from the theoretical standpoints. A generalized Rayleigh–Plesset equation has been developed that includes the effects of a thick viscous fluid shell. The equation is compared with other theoretical encapsulated bubble formulations and some recent experimental observations. For a 4.0 μ bubble, with a 500 nm shell of viscosity of 28 cP, insonified over 5 cycles, a maximum relative expansion of 2.2 times the equilibrium radius is predicted at 2.25 MHz and 3.1 times at 1.5 MHz. Predictions from this model regarding the scaling of radial expansion with changes in pressure and frequency have been validated with experimental data.

11:30

1aBB8. Quantification of submicron DDFP gas bubbles using 30 MHz backscattered ultrasound. E. Carr Everbach (Eng. Dept., Swarthmore College, Swarthmore, PA 19081), Ronald A. Roy (Boston Univ. Mech./Aerosp. Dept.), and John T. Flaherty (Sonus Pharm., Inc.)

To determine the size distribution of microbubbles following a temperature-induced phase change of dodecafluoropentane (DDFP) emulsion particles, the acoustic backscatter was measured in 2 μs duration tonebursts of 30 MHz ultrasound from individual bubbles present in the emulsion as it moved in a coaxial jet flow at 20 °C and 37 °C [R. A. Roy and R. E. Apfel, ''Mechanical characterization of microparticles by scattered ultrasound,'' J. Acoust. Soc. Am. **87**, 2332–2341 (1990)]. DDFP liquid emulsion droplets would be expected to undergo a phase change to gaseous microbubbles above the DDFP boiling point of 28–30 °C at atmospheric pressure. Calibration of the backscatter coefficients for the system was accomplished using 6 micron diameter polystyrene spheres. The DDFP bubbles were assumed to be small enough to behave isothermally and viscosity−dominated in their dynamics [D. Khismatullin and A. Nadim, ''Radial oscillations of encapsulated microbubbles in viscoelastic liquids,'' Phys. Fluids, submitted]. The results yielded a bubble size distribution for the emulsion of between 50 and 100 nm radius with a mean radius of 75 nm at 37 °C. [Work supported by a grant from Sonus Pharmaceuticals, Inc.]

11:45

1aBB9. Analysis of deformation process of cells and internal bubbles by shock waves for developing DDS and bioprocess. Masaaki Tamagawa, Atsushi Matsumoto (Grad. School of Energy Sci., Kyoto Univ., Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan, tama@energy.kyoto-u.ac.jp), and Ichiroh Yamanoi (P&G Far East, Inc., Kobe 658-0032, Japan)

This paper describes the fundamental investigation for developing drug delivery systems (DDS) by shock waves and high efficient environmental bioprocess by shock wave, in which the microcapsules and cells have internal bubbles to crack the membrane easily. In this study, the analysis of flow using computational fluid dynamics (CFD) and optical visualizations for these two common phenomena were carried out. As for the analysis, the effects of the differences of acoustic impedance between internal fluid media and the outer cell and moving boundary of the cell on the pressure were examined. The result showed that the effects were large for the deformation process of the internal bubble and cell. Regarding the visualization, the deformation process of a bubble near the elastic wall as a model of cell membrane by shock waves was observed to analyze the behavior of the bubble in the cell for bioprocess or microcapsule for DDS. From these observations of deformation process, it was suggested that there should be optimized or critical point to have large deformation and displacement of the bubble by changing the initial position and diameter of the bubble in the cell and relative curvature of the elastic wall (cell diameter).

MONDAY MORNING, 3 DECEMBER 2001 ROOM 305, 8:00 TO 10:55 A.M.

Session 1aSPa

Signal Processing in Acoustics: Topical Meeting on Acoustic Time Reversal and Applications: Theory and Basic Experiments in Time Reversal

David H. Chambers, Chair

Electrical Engineering, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, California 94551-5508

Chair's Introduction—8:00

Invited Papers

8:05

1aSPa1. An overview of time-reversed acoustics. Mathias Fink (Laboratoire Ondes et Acoustique, ESPCI, Universite Denis Diderot, Paris, France)

Time-reversal invariance is a very powerful concept in classical and quantum mechanics. In the field of acoustics, where time-reversal invariance also occurs, time-reversal experiments may be achieved simply with arrays of transmit–receive transducers, allowing an incident acoustic field to be sampled, recorded, time-reversed, and re-emitted. The time-reversal mirrors (TRMs) are innovative tools in the field of fundamental physics. They may be used to study random media, multiple scattering processes, chaotic scattering, inverse scattering problems, dissipation effects, and diffraction limits. They open the way to new signal processing. An overview of these fields will be presented. Applications of TRMs will also be described, including underwater acoustics, medical applications, telecommunications, as well as nondestructive testing.

9:20

1aSPa2. Ultrasonic time reversal, basic principles. Didier Cassereau and Mathias Fink (Laboratoire Ondes et Acoustique, 10 Rue Vauquelin, 75005 Paris, France)

For about 10 years now, time reversal methods have been widely developed in the laboratory and applied to various domains of acoustics like medical imaging/therapy and nondestructive evaluation. Many experiments and theoretical works have shown the robustness of this approach in multiple conditions (inhomogeneous and multiscattering media, solid layers, submarine acoustics, . . .). This paper will first present an overview of the basic theoretical principles which govern the time reversal approach. It will be shown how the concept of closed time reversal cavity yields theoretical predictions of the intrinsic limitations of the process. Starting from the closed time reversal cavity, the discussion will migrate in direction of (i) time reversal mirrors and (ii) time reversal arrays of final aperture. It will be shown how the general performances of the system are degraded. Then the presentation will focus on the time

reversal methods applied to ultrasonic waves in solids. It will be shown how the time reversal technology takes advantage of longitudinal and transverse propagation modes in order to increase the focusing efficiency. These results will be illustrated by an experiment carried out with an aluminum plate partially immersed in water and excited by a laser impact.

9:40

1aSPa3. An overview of the D.O.R.T. method. Claire Prada (Laboratoire Ondes et Acoustique, ESPCI, Université Paris 7, U.M.R. 7587, 10 rue Vauquelin, 75005 Paris, France, claire.prada-julia@espci.fr)

This talk will review the decomposition of the time reversal operator (D.O.R.T.) method. This new approach to active detection and focusing of acoustic waves using arrays of transmitters and receivers allows selecting and focusing on scatterers through complex media. The method was derived from the theoretical analysis of iterative time reversal mirrors. It consists essentially in the construction of the invariants of the time reversal process. The diagonalization of the time reversal operator provides both a measure of the strength of a scatterer and the phase law that can be applied to the array to focus selectively on a scatterer. After explaining the principle of the D.O.R.T. method several experimental results will be shown.

10:00–10:15 Break

10:15

1aSPa4. Spectrum of the time reversal operator. David Chambers (Lawrence Livermore Lab., P.O. Box 808 L-154, Livermore, CA 94551) and A. K. Gautesen (Dept. of Mathematics, 136 Wilhelm Hall, IA State Univ., Ames, IA 50011)

In a series of articles beginning in 1994, Prada and Fink showed that the behavior of any time reversal system could be determined by an analysis of a time reversal operator. This operator incorporates the effects of geometry, the background medium, and the scattering characteristics of objects within the medium. Initial analysis and experimental measurements for small, neutrally buoyant, point-like objects in a homogeneous medium showed a one-to-one correspondence between the eigenstates of the operator and individual scatterers. Recent work by Chambers and Gautesen for small spheres with arbitrary densities showed multiple eigenstates could exist for a single scatterer, breaking the one-to-one correspondence. In this talk, the spectral analysis of the time reversal operator is reviewed and new results are shown for compact scatterers of finite size and for an extended scatterer (a wire). It is shown that the number of eigenvalues for a given scatterer can be indefinitely large depending on the scattering physics. Furthermore, the spectrum for a collection of noninteracting objects is the combination of individual spectra from each scatterer. [Work performed under the auspices of the Department of Energy by the Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.]

10:35

1aSPa5. Nonlinear acoustics of phase conjugate ultrasound waves. A. P. Brysev, L. M. Krutyansky (Wave Res. Ctr., General Phys. Inst., Russian Acad. of Sci., 38 Vavilov St., 119991 Moscow, Russia), M. F. Hamilton (Univ. of Texas, Austin, TX 78712-1063), P. Pernod, and V. I. Preobrazhensky (Ecole Centrale de Lille, 59651 Villeneuve d'Ascq Cedex, France)

Nonlinear propagation of phase conjugate waves (PCW) in liquid is studied experimentally and simulated numerically. A supercritical parametric technique is used for ultrasonic (5–10 MHz) wave phase conjugation (WPC) with enormous amplification (more than 80 dB). The technique provides sufficient PCW intensity for vivid manifestation of acoustic nonlinearity. Quantitative studies of refocused PCW beams have been performed. The space–time structure of nonlinear PCW was measured and results were compared with calculations. Finite aperture size and conjugator gain were taken into account in a numerical model based on the nonlinear parabolic wave equation. Accurate refocusing of strongly nonlinear PCW beams is demonstrated for sufficiently large apertures. The second and higher harmonics generated by PCW are localized within the focal region of the incident wave even after PCW refraction in a randomly inhomogeneous medium. The phenomenon is explained theoretically by phase locking of PCW harmonics in nondispersive acoustic media. The results have been generalized to include narrow-band WPC of selected harmonics of the incident wave when the latter is nonlinear. Specifically, WPC can both remove phase distortion and amplify the second harmonic sufficiently that imaging can be performed with the resulting fourth harmonic. This novel imaging technique is demonstrated in an example of C-scan acoustic microscopy.

Session 1aSPb

Signal Processing in Acoustics: Topical Meeting on Acoustic Time Reversal and Applications: Time Reversal in Signal Processing

James V. Candy, Chair
Lawrence Livermore National Laboratory, P.O. Box 808, L-156, Livermore, California 94551

Invited Papers

10:55

1aSPb1. Time-Reversal signal processing: An overview. James V. Candy (Lawrence Livermore Natl. Lab., Univ. of California, P.O. Box 808, L-156, Livermore, CA 94551)

The development of time reversal (T/R) techniques both theoretically and pragmatically has heightened in interest over the last decade. Although the majority of the effort has been concentrated in analysis and proof-in-principle experiments, the evolution of these results have led to potential applications in signal processing. These range from the detection of flaws in nondestructive evaluation (NDE) and scatterers in ocean acoustics to the destruction of kidney stones in biomedical ultrasound. When the medium under investigation has quasihomogenous properties, imaging using the T/R operator has evolved as a viable entity. In this paper we present an overview of T/R from a signal processing perspective. Here we concentrate on the ability of the T/R to perform its primary function of ''focusing'' on the strongest scatterer. We show how the T/R properties provides us with the optimal spatio-temporal matched-filter property which is utilized in a wide variety of applications, especially in NDE, biomedical, and the newly investigated communications area. We note the relationship of T/R to the inverse solution providing an alternative in deconvolution-type problems.

11:15

1aSPb2. DORT applied to active sonar. David M. Fromm and Charles F. Gaumond (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375-5320)

The decomposition of the time reversal operator (DORT) is a single frequency method of spatially isolating scatterers with a multiple-source/multiple-receiver system [C. Prada *et al.*, J. Acoust. Soc. Am. **99**, 2067–2076 (1996)]. The application of the DORT technique to a broadband sonar system is explained by showing its relationship to the sonar equation. The orthogonal matrices are related to source and receiving steering vectors and the singular values to the echo levels of the various signal components. The structure of a coherent, reverberation-limited, broadband, numerical sonar model is shown. Numerical results from this sonar model are processed by windowing and Fourier transforming. Then DORT is applied and the depth isolated target components are identified through projecting source and receiver steering vectors with modeled propagation matrices. The ability of this procedure to isolate scatterers from distributed bottom reverberation is shown for a set of source/receiver geometries with the scatterer at various depths and ranges. Both range-independent and range-dependent environments are considered. [Research sponsored by ONR.]

11:35

1aSPb3. Computational time-reversal and object localization. Anthony J. Devaney (Dept. of Elec. and Computer Eng., Northeastern Univ., Boston, MA 02115)

The problem of detecting and locating point scatterers embedded in a known, nonrandom, background medium from multi-static scattering data specified at one or more temporal frequencies is addressed within the context of the acoustic wave equation. It is shown that the singular value decomposition (SVD) of the multi-static data matrix leads to a general theory for both target detection and location that applies for arbitrary distributions of sources and receivers and that reduces to the standard method of computational time-reversal in the special case where the sources and receivers are co-located. The developed theory employs both standard backpropagation as well as vector subspace methods such as MUSIC to perform target location estimation and is shown to apply to such diverse applications as ultrasound imaging, acoustic wave tomography, and offset vertical seismic profiling (VSP). The talk includes computer simulations as well as results from experimental data.

11:55

1aSPb4. Time-reversal acoustics and maximum-entropy imaging. James G. Berryman (Lawrence Livermore Natl. Labs., P.O. Box 808 L-200, Livermore, CA 94551-9900)

A common problem in acoustics imaging is target location using either passive or active data inversion. Time-reversal methods in acoustics have the important characteristic that they provide a means of determining the eigenfunctions and eigenvalues of the scattering operator for either of these two cases. Each eigenfunction may often be approximately associated with an individual scatterer. The resulting decoupling of the scattered field from a collection of targets is a very useful aid to localizing the targets, and suggests a number of imaging and localization algorithms. Two of these are linear subspace methods and maximum-entropy imaging.

Session 1aUW

Underwater Acoustics: Propagation Modeling and Environmental Effects I

Paul A. Baxley, Chair
Space and Naval Warfare Systems Center, Acoustics Branch, 49575 Gate Road, San Diego, California 91252-6435

Chair's Introduction—8:25

Contributed Papers

8:30

1aUW1. Acoustic wavefront modeling in shallow water. Chris T. Tindle (Phys. Dept., Univ. of Auckland, Auckland, New Zealand)

Wavefront modeling offers a direct method of finding pulse waveforms in shallow water and allows for both ocean surface waves and variable water depth. The wavefront is found by tracing rays to a fixed range and finding the ray depth as a function of starting angle. Interpolation on the wavefront provides the amplitude and arrival time at the receiver depth. Phase is determined by the ray history. Surface waves can lead to focusing with cusps and caustics in the wavefront. Near cusps and caustics the field is found by numerically fitting a phase function with the appropriate number of saddle points. Model results are compared with field measurements. [Work supported by ONR.]

8:45

1aUW2. Three-dimensional Gaussian beam tracing for shallow-water applications. Paul A. Baxley, Homer Bucker, Vincent K. McDonald (Space and Naval Warfare Systems Ctr., San Diego, Acoust. Branch, D857, 49575 Gate Rd., Rm. 170, San Diego, CA 91252-6435, baxley@spawar.navy.mil), and Michael B. Porter (Sci. Applications Intl. Corp., Ocean Sci. Div., 888 Prospect St., Ste. 201, La Jolla, CA 92037)

Gaussian beam tracing is a ray-based method of acoustic wave propagation that overcomes some of the implementation problems of conventional ray methods, while retaining a speed advantage over wave theory approaches. A fan of Gaussian beams are propagated from a source, according to the standard ray equations, so that the pressure at any given point can be approximated by the sum of contributions from each beam. Eigenray computations, perfect shadows, and infinite caustics associated with standard ray methods are thereby eliminated, and the applicability to lower-frequency problems is improved. Gaussian beam tracing is particularly well suited for three-dimensional problems, for which eigenray computations are costly. The effectiveness of a three-dimensional Gaussian beam model [Bucker, J. Acoust. Soc. Am. **95**, 2437–2440 (1994)] for predicting acoustic propagation in shallow water is demonstrated. At low frequencies, Gaussian beam and wave solutions are compared for the ASA benchmark wedge problem and the Shallow Water Acoustic Modeling 1999 Workshop Gaussian canyon problem. The use of the model for the analysis of three-dimensional bathymetric effects in real data will also be demonstrated. At high frequency, model predictions of impulse responses will be compared to measured data for a number of acoustic communications experiments. [Work supported by ONR.]

9:00

1aUW3. Analysis of the effects of numerical dispersion on pulses in finite-difference and pseudospectral time-domain methods. Gareth R. Elston, Judith M. Bell, and Chris G. Capus (Dept. of Computer & Elec. Eng., Heriot-Watt Univ., Riccarton, Edinburgh EH14 4AS, UK)

Finite difference (FD) and pseudospectral (PS) time-domain methods are increasingly used for underwater acoustics modeling. Numerical dispersion is unavoidable but can be reduced with an appropriate choice of temporal and spatial discretization (Δt and Δx). As pulses propagate, numerical dispersion causes time-dilation which increases with simulation time and pulse bandwidth. Analysis of the dispersion relation leads to an equation which shows that the time-dilation in a large FD model can be significant for well defined pulses of 50 points per wavelength (PPW). Reducing the dispersion requires a higher grid density. For a 2D model, the memory requirement and time costs vary as the inverse square and inverse cube of Δx. This is contrasted with the PS model in which the time-dilation can be set arbitrarily small for any choice of Δx by reducing Δt. For a fixed spatial grid density, e.g., 10 PPW, the computational cost is linear in the temporal grid density. A locally optimized short-time fractional Fourier transform shows that numerical dispersion has a dramatic effect on the time-frequency characteristics of dispersed pulses, e.g., an input CW pulse becomes a nonlinear upward chirp. These effects are illustrated for a variety of pulses.

9:15

1aUW4. Nonadiabatic travel time and time series synthesis in range dependent waveguides. Kevin D. LePage (SACLANT Undersea Res. Ctr., Viale San Bartolomeo 400, 19138 La Spezia, Italy, lepage@saclantc.nato.int)

A WKB approximation for the range dependent modal group slownesses and dispersions in a range dependent waveguide is derived. The resulting slownesses can be used to synthesize time series in a narrow band or over broader bandwidths by adding narrow band results together coherently. The slownesses themselves are also highly useful intermediate products which help give insight into the speeds with which the modes of a range dependent waveguide propagate. The range dependent modal group slownesses are compared to their adiabatic approximations and conclusions are drawn regarding the relative accuracy of these two approximations for time series synthesis.

9:30

1aUW5. A comparison of techniques for simulating the signal received by a moving hydrophone at short range in shallow water. Alec J. Duncan (Ctr. for Marine Sci. and Technol., Curtin Univ., GPO Box U1987, Perth, WA 6845, Australia, a.duncan@cmst.curtin.edu.au)

This paper presents two techniques for time series simulation applicable to the case of a moving hydrophone a few water depths from a broadband sound source. The first technique is based on the well-known image method but extends its applicability to low frequencies by accurately incorporating the physics of the seabed reflection process. The second technique is based on Fourier synthesis using transfer functions calculated using a standard fast-field propagation code. Transient and continuous-wave signals calculated using the two methods are shown to be in excellent agreement, even at frequencies where the water depth is less than half a wavelength. [Work supported by the Defence Science and Technology Organisation, Australia (DSTO).]

9:45

1aUW6. Propagation through an imperfectly layered fine-structure. John J. McCoy (The Catholic Univ. of America, Washington, DC 20064) and Ben Z. Steinberg (Tel Aviv Univ.)

The ''fine-structure'' refers to spatial variations that obtain on a length scale, l, that is small when compared to the acoustic wavelength, λ, in a reference medium, i.e., $\lambda \gg l$. The ''imperfection'' refers to variations in *measures* of the fine structure, which obtain on a much larger length scale, L. A precise definition of a useful ''global'' measure of the fine-structure variation is determined by the *mathematical model* of the experiment described. The variations that apply to this measure are required to be *weakly varying* and *stochastic*. Prediction models can be expressed in *statistical moments* of a *spatially smoothed component* of the complete propagating wavefield; the first two moments are considered. The model constructions are accomplished in two steps, the first of which is a ''generalization'' of classical homogenization theory, now based on a *multiresolution decomposition using wavelets*. The second step is to derive equations that govern the statistical moments of solutions of the homogenized equations. An alternate construction, for which an equation that determines the statistical moments of the complete acoustic field is obtained before accomplishing the homogenization, is discussed.

10:00–10:15 Break

10:15

1aUW7. A rational approximation for the parabolic equation method. Ya Yan Lu (City Univ. of Hong Kong, Kowloon, Hong Kong)

The parabolic equation method typically uses rational approximations to the square root operator. The split-step Padé method of Collins is more efficient and it uses rational approximations of the propagator—the exponential of the square root operator, say $P=\exp is\sqrt{1+X}$. Previously, only approximants of the type $[p/p]$ are used. In order to damp the evanescent modes (and avoid instabilities), these $[p/p]$ approximants are not the simple $[p/p]$ Padé approximant of P. In this talk, we demonstrate that $[p/(p+1)]$ Padé approximant of P can be used in the split-step Padé method. It is simple to construct and it has a good accuracy for approximating the propagating modes, yet it is guaranteed to suppress the evanescent modes. This approach will be compared with other variants of the split-step Padé method in numerical examples for acoustical waveguides.

10:30

1aUW8. Joint channel and Doppler estimation via low-complexity prior. Paul Gendron and T. C. Yang (Naval Res. Lab., 4555 Overlook Ave., Washington, DC 20375)

Joint estimation of channel impulse response and Doppler drift can be confounding since source motion and media temporal fluctuations can be misidentified. In addition, multipaths can be associated with different though not independent Dopplers. We present a novel algorithm for joint estimation of channel impulse response and mean Doppler drift by assuming prior constraints that favor lower complexity of the posterior model. The resulting estimator is compared with the popular maximum likelihood estimator (MLE). Approximate posterior covariance estimates provide Doppler/channel correlation and confidence bounds. We test this method with synthetic channel data and experimental data from a shallow water environment and compare the method with the classical MLE.

10:45

1aUW9. Characterization of temporal fluctuations of broadband channel impulse functions within communication packets. T. C. Yang and Paul Gendron (Naval Res. Lab., Code 7120, 4555 Overlook Ave., Washington, DC 20375)

Temporal fluctuations of an acoustic channel can be characterized by the amplitude and phase fluctuation statistics of individual multipath arrivals, and the temporal coherence of the channel. At low frequencies where the temporal coherence time is much longer than the duration of the probing pulse, the measurements of the fluctuation statistics and channel coherence functions were straightforward and can be found in previous publications. As frequency increases to tens of kHz or above, the coherence time can be shorter than the packet length used in acoustic communications (ACOMMS). Under such circumstances, the channel coherence can impede the equalizer performance and measurement of the temporal evolution of the channel within the duration of a packet (the intra-packet coherence) is critical for modeling the ACOMMS performance. In this paper we report the measurements of intra-packet coherence and multipath fluctuation statistics from two at-sea experiments using M-sequence signals having sequence length much shorter than the ACOMMS packet length. Initial analysis results associated with fixed–fixed and moving-drifting source–receiver configurations will be presented. [Work supported by ONR.]

11:00

1aUW10. Performance metrics for underwater communications: An initial assessment. Michael S. Richman, Ned B. Thammakhoune, Scott L. Whitney, and Geoffrey S. Edelson (BAE Systems, M/S MER15-2651, P.O. Box 868, Nashua, NH 03061-0868)

The relationship between the performance of an underwater acoustic data communications system and operating conditions is investigated. There has been much work devoted toward this end focused on developing mathematical models of the underwater acoustic channel. However, the underwater acoustic environment is a complicated one, and current models do not fully account for the challenging, time-varying conditions. The analysis here is based on actual data transmissions collected during several at-sea tests involving a variety of operating conditions. Metrics have been developed that attempt to characterize properties relating to signal-to-noise ratio, Doppler, multipath spread, channel coherence, and channel complexity. By applying these metrics to data transmissions in conditions with different water depths, carrier frequencies, and source/receiver geometries, a better understanding can be developed of the factors that determine the performance of an underwater communications system. The ultimate goal of this work is to provide a systematic prediction method by computing those metrics that are the true arbiters of performance. This work will also aid in the development and validation of mathematical models of the underwater acoustic environment by providing practical measures with which to compare the simulated channels.

11:15

1aUW11. Treatment of ocean bottoms with a layered bottom with some layers having a lower impedance than some overlying layers. M. F. Werby (NRL Code 7180, Stennis Space Ctr., MS 39529 and Dept. of Phys., Catholic Univ. of America, Washington, DC 20064)

Some oceans appear to have layers that are less dense and have slower speeds than overlying layers. Perhaps this is due to a gaseous admixture with some of the sediment. This presents some interesting effects in model calculations which are dependent on frequency, layer thickness, and adjacent layers. The mathematical consequence of this is examined.

Session 1pAA

Architectural Acoustics: Room Acoustics Potpourri

David E. Marsh, Chair
Pelton Marsh Kinsella, 1420 West Mockingbird Lane, Suite 400, Dallas, Texas 75247

Chair's Introduction—1:30

Contributed Papers

1:35

1pAA1. Concert hall acoustics—2001. Leo L. Beranek (Acoust. Design Consultant, 975 Memorial Dr., Cambridge, MA 02138-5755)

A brief summary is made of major factors encountered in today's acoustical design of concert halls, including preferred reverberation times, loudness, lateral reflections, seat absorption, suspended panels, need for irregular surfaces to create sound diffusion and eliminate acoustical ''glare,'' and objective measurement techniques developed from laboratory experiments that can be used to estimate subjective acoustical quality. Objective measurements of these factors have been made in over 70 halls. The results are plotted versus subjective rank-orderings of acoustical quality for these halls. The rank-orderings were obtained from interviews and questionnaires involving conductors, music critics, and music aficionados. The objective quantities plotted are reverberation time, early decay time, spaciousness, lateral fraction, interaural cross-correlation coefficient, initial-time-delay gap, strength factor (sound level), bass ratio, 125-Hz level, and surface diffusivity index.

1:50

1pAA2. Subjective and objective evaluations for solo and chamber music performances. Weihwa Chiang (Taiwan Univ. of Science and Technol., 43, Section 4, Kee-lung Rd., Taipei City, Taiwan) and Weichun Wang (Saint Mary and Saint Jones Inst. of Technol., Dansui, Taipei County, Taiwan)

Stage acoustics were evaluated in six halls regarding solos (piano, harp, and violin) and chamber music (piano sonata, piano trio, and brass quintet). In each hall professional musicians were seated at two locations: with and without detached side reflectors. The performers expressed their opinions about hearing oneself, hearing others, ease of ensemble, and overall impression. Subjects in the audience were asked about the strength of individual parts and overall impression. A factor analysis yielded four groups among the musicians. Regarding overall impression string players favored the environments with side reflectors. An early-to-direct energy ratio (ED100) was more important than early support (STearly) and late support (STlate) subjectively. The optimum value of ED100 was V11 dB. The overall impression evaluated by the musicians was poorly correlated with the balance of individual parts evaluated by the audiences, indicating that there were orthogonal factors for designing the stage for solo and chamber music performances. [Work supported by the National Science Council of Taiwan.]

2:05

1pAA3. The assessment of room acoustics parameters using the MLS technique and numerical simulation. Samir N. Y. Gerges and Marcio H. A. Gomes (Universidade Federal de Santa Catarina (UFSC-CTC-EMC), Trindade, Florianopolis, SC, Brazil, CEP:88040-900)

Advances in psychoacoustics through the years, have made it possible to evaluate the acoustical quality of a room, based on several numerical parameters that have been developed. Even though there is not a total agreement about which parameters are truly important, some of them are accepted by most of the acousticians. When designing a concert hall, ray-tracing software can be used to calculate some of these parameters. However, some input data carry some uncertainties. Available data for absorption coefficients have to be used carefully and, beyond that, there is not a standardized recommendation for measuring diffusion coefficients, nor a data base for this parameter. This paper presents a comparison of a room simulated with one of the commercial software available (RAYNOISE), and the measured parameters in this same room, using maximum length sequence (MLS). Discrepancies over the results are investigated and analyzed.

2:20

1pAA4. Overhead localization in a hall with oval plan, and its practical solution. Toshiyuki Okano (Takenaka Res. & Development Inst., 1-5-1, Ohtsuka, Inzai, Chiba 270-1395, Japan, okano.toshiyuki@takenaka.co.jp)

This report concerns an unusual localization phenomenon that was observed on the center line of a hall having an oval plan. The hall, recently completed in Tokyo with 760 seat capacity and one balcony, is used primarily for chamber music. After installation of audience seats, a troublesome overhead localization was observed at listener positions on the fore–aft center line whenever the source was also on the center line. Listening tests at the site indicated that the locations from which the unwanted reflections appeared to originate were the following parts: (a) two opposite portions of the concave side walls just below the ceiling; (b) main-floor rear side walls; and (c) a portion of the main-floor rear wall. It was demonstrated that temporarily covering those surfaces with sound absorbing material could drastically reduce the overhead localization. In order to avoid seriously decreasing the reverberation time, a minimum amount of added sound absorption was a necessity. Tests in the laboratory demonstrated that the absorbing areas on part (b) would be minimized if installed asymmetrically. The final result rendered the localization phenomenon and some on-stage rear-wall echo inaudible when sound absorbing materials were permanently applied.

2:35

1pAA5. How worship space acoustics helped to shape western history. David Lubman (David Lubman & Assoc., 14301 Middletown Ln., Westminster, CA 92683)

Acoustical communication is a vital element in the co-evolution of worship space architecture and liturgy. Acoustical communication in worship spaces is bound by scientific verities. Information theory holds that for any communication channel there exists an optimum coding of transmitted information. Speech and music are alternative informational codings. Speech coding is efficient for small, dry indoor spaces and well suited to the spoken liturgies of synagogues, early Christian house

churches, and other small pre-Constantinian churches. The historic and fateful change from a spoken to a musical liturgy beginning with Constantine (ca. 313) was likely driven by the highly reverberant acoustics of large basilicas that rendered speech unintelligible. Such highly reverberant spaces are well suited for specific types of musical communication. Gregorian chant is shown to be an ingenious coding for reverberant spaces and an early use of frequency diversity coding. The historic return to spoken liturgies beginning with the reformation and counter-reformation of the 15th century was accompanied by reductions in church reverberation. The Protestant acoustic also engendered a need for music suited to less reverberant spaces, e.g., simple melodies with rapidly changing chordal harmonies. Electronics is a new element driving the evolution of worship space architecture and liturgical form.

2:50–3:00 Break

3:00

1pAA6. A revision of the accepted criteria for the distribution of the resonance frequencies of rectangular rooms. Richard A. Moscoso (Seccion Fisica, Pontificia Universidad Catolica del Peru, Apartado 1761, Lima, Peru) and Caleb Vasquez (Centro de Producciones de la Union Peruana, Av. Comandante Espinar 670, Miraflores, Lima, Peru)

There are several criteria to predict to good or bad mode distribution in function of the proportion between the dimensions of a rectangular enclosure. The principals are the criterion of Bolt and the criterion of Bonello. Both, do not demand a fix relationship between the proportions of an enclosure, but rather permit a wide range of combinations between them. To generate an adequate distribution of modes so that the coloration will be void or at least be the minimal possible is necessary to take in account the reverberation time. The criterion of Bolt is independent of the enclosure reverberation time. The criterion of Bonello has an improvement respect to the criterion of Bolt because it is sensible to the volume changes, but it is not directly related to the reverberation time. The present work shows the procedure to obtain the theoretical reverberation time to avoid coloration in a rectangular enclosure, analyzing the modal spectral density in 1/3 octave bands. This is critical in studio control rooms and studio recording rooms.

3:15

1pAA7. Complex resonance frequencies numerically obtained in the rectangular room with arbitrary wall impedances. Sylvio R. Bistafa (Dept. of Mech. Eng., Polytechnic School, Univ. of Sao Paulo, Sao Paulo, 05508-900, SP, Brazil) and John W. Morrissey (Numerical Algorithms Group—NAG, Inc., Downers Grove, IL 60515-5702)

Numerical methods to finding the complex resonance frequencies in the rectangular room with arbitrary wall impedances are presented. The eigenvalue equation is written in such a way that allows the unambiguous identification of each Riemann sheet. One numerical procedure applies the Newton method. For each wall-pair mode, the complex resonance frequencies are found, in each Riemann sheet, by small perturbations to the impedances of the wall pair. Another procedure transforms the eigenvalue equation into two differential equations, which are numerically integrated by the homotopic continuation procedure. The latter procedure is faster and finds all possible solutions. From the knowledge of sound decays experimentally obtained in a hard walled rectangular room, the impedances of the hard walls are obtained, which are then used to predict sound decays in the same room with one surface lined with sound absorbing material. It is shown that groups of modes with different reverberation times may collectively produce sound decays not always consistent with the Sabine prediction, and with some experimental results in a rectangular room with the floor covered with sound absorbing material of known acoustic impedance.

3:30

1pAA8. Prediction of the characteristic of sound fields in long enclosures by a hybrid computer model. Xiangyang Zeng, Kean Chen, and Jincai Sun (Post Box 19, Northwestern Polytech. Univ., Xi'an 710072, PROC)

A combined ray tracing and image source model RATISM has been developed specially for the simulation of sound propagation throughout long enclosures such as a corridor. How the algorithm models sound source, sound propagation, and sound receiver was described. The ways in which more complex long enclosures can be modeled were also given. The methods of the prediction of reverberation time and sound pressure level were presented in detail. Three different enclosures were modeled and the results were analyzed. The reverberation room of Northwestern Polytechnic University was measured and predicted to show RATISM can be used for both diffused and nondiffused sound field. A rectangular long enclosure was then predicted by RATISM and NORMAL [L. N. Yang and B. M. Shield, J. Sound Vib. **229**, 133–146 (2000)] and the comparison showed that both the two models can correctly simulate the sound field in rectangular long enclosures. A more complicated long enclosure was also simulated and the results showed that the sound field in it had similar characteristic to that in the rectangular one, in which the sound field is nondiffused and sound distribution was mainly dependent on the distance of receiver to source. [Work supported by the Doctorate Foundation of Northwestern Polytechnical University.]

3:45

1pAA9. A combined auralization and visualization system for noise prediction in enclosures. Xiangyang Zeng, Fangli Ning, Ke'an Chen, and Jincai Sun (Post Box 19, Northwestern Polytech. Univ., Xi'an 710072, PROC)

A combined auralization and visualization system was devised for the research on internoise control. For any individual point in an enclosure, a hybrid geometric method was used to simulate the impulse response at high frequency and the finite element method was used to model that at low frequency. Then the two parts were joined. Acoustical indexes such as SPL(A) can be calculated based on this complete impulse response. For binaural simulation, the impulse responses of two points, each of which was 0.08 m away from the receiving point were considered as the binaural impulse response. They were then convoluted with any anechoic signal in time domain using the fast Fourier transform algorithm. This subsystem of auralization was tested on real acoustical spaces to compare real sound signals and the corresponding reproduced ones. The subsystem of visualization was designed to build an interface for data input and output and to show the structure of the enclosure and status of sound sources and receivers. By using computer language C++ these two subsystems were combined, therefore the complete system can be conveniently used on a personal computer. [Project supported by the Doctorate Foundation of Northwestern Polytechnical University.]

Session 1pBB

Biomedical Ultrasound/Bioresponse to Vibration: Propagation and General Topics

Ibrahim M. Hallaj, Chair
Wolf, Greenfield and Sacks, PC, 600 Atlantic Avenue, Boston, Massachusetts 02210

Chair's Introduction—1:00

Contributed Papers

1:05

1pBB1. Measurement of guided ultrasonic waves in the human tibia and radius. Patrick H. Nicholson, Sulin Cheng (Dept. of Health Sci., Univ. of Jyvaskyla, 40351 Jyvaskyla, Finland), Tommi Karkkainen (Dept. of Mathematical Information Technol., Univ. of Jyvaskyla, 40351 Jyvaskyla, Finland), Petro Moilanen, and Jussi Timonen (Dept. of Phys., Univ. of Jyvaskyla, 40351 Jyvaskyla, Finland)

The use of guided ultrasonic waves to characterize human long bones may offer advantages over existing approaches based on lateral wave velocity measurement alone. To investigate this possibility, a pulse transmission system consisting of low frequency transducers orientated perpendicularly to the limb was used to measure the tibia and radius of healthy and osteoporotic subjects. At least two wave components were consistently observed: a fast wave traveling at 3600–4500 m/s, and a slower wave propagating at 1200–1900 m/s. Satisfactory excitation of the slower mode was dependent on adequate transducer contact pressure. Analytical and numerical models, combined with phantom measurements, were used to confirm the nature of the two waves. The fast wave behaved as expected for a lateral wave, while the slower wave was consistent with the lowest order antisymmetric guided mode. In pilot studies, the relative velocity difference in the tibia between healthy and osteoporotic women was greater for the guided wave (18%) compared to the lateral wave (2%), suggesting that the guided wave is a better discriminant of osteoporosis. These preliminary results indicate guided wave analysis to be a promising adjunct to conventional axial ultrasound measurements in bone.

1:20

1pBB2. Scattering-induced attenuation of a fast longitudinal wave propagating in a trabecular bone modeled as a nondissipative porous medium. Francine Luppe, Jean-Marc Conoir, and Herve Franklin (LAUE, UMR CNRS 6068, Universite du Havre, Pl. R. Schuman, 76610 Le Havre, France)

In this work, the calcaneus is modeled as a porous medium, made up with cortical bone and marrow. The Biot theory is used to describe acoustic propagation, in the 0.2–1 MHz range. In this frequency domain, the marrow viscosity effect is negligible, so that the calcaneus is considered as a nondissipative porous medium. In that same frequency range, however, the largest pores can not be included in the Biot modelization, as they may be larger than the acoustic wavelengths. They are hence supposed to act as scatterers, which radiate energy in other directions than an incident fast longitudinal plane wave. By use of an independent scattering approximation, with a unit scattering cell composed of two marrow filled cylinders, the effective attenuation induced by multiple scattering is studied, for different values of bone porosity and spatial arrangement of the two cylindrical pores considered as one single scatterer.

1:35

1pBB3. Investigation of guided ultrasonic waves in bone phantoms using an axial transmission method. Petro Moilanen, Jussi Timonen (Dept. of Phys., Univ. of Jyvaskyla, PL35, 40351 Jyvaskyla, Finland), Tommi Karkkainen, Patrick H. Nicholson, and Sulin Cheng (Dept. of Health Sci., Univ. of Jyvaskyla, PL35, 40351 Jyvaskyla, Finland)

Human long bones are effectively hollow tubes of irregular cross-section which potentially allow the propagation of guided modes in addition to the lateral or ''head'' wave measured by existing clinical devices. Guided modes are attractive because they propagate throughout the bone thickness, are less attenuated with distance, and may yield a more complete picture of the bone material and geometrical properties. A low-frequency pulse transmission technique was used in bone phantoms consisting of acrylic plates with thicknesses ranging from 2 to 25 mm. The receiving transducer was moved automatically, producing (r,t) plots which allowed different wave fronts to be visualized and measured. Two distinct wave components were observed. The faster wave front behaved as predicted for the lateral wave, propagating at the longitudinal velocity (2700 m/s) in thick plates. However, in plates thinner than 5 mm this velocity decreased nonlinearly with thickness toward that of the lowest-order symmetric guided wave (2400 m/s). The second component was consistent with the lowest-order antisymmetric guided wave, propagating at the Rayleigh velocity (1300 m/s) in the thickest plates and exhibiting a strong nonlinear velocity decrease at thicknesses below 10 mm. These results provide a useful framework for applying similar measurements in real bones.

1:50–2:05 Break

2:05

1pBB4. Estimation of the Doppler shift frequency using selected phase information for high-frame rate blood flow imaging. Naohiko Tanaka (Dept. of Systems Eng., Shibaura Inst. of Technol., Saitama 330-8570, Japan, tanaka@se.shibaura-it.ac.jp) and Shigeo Ohtsuki (Tokyo Inst. of Technol., Yokohama 226-8503, Japan)

A new signal processing approach is presented for the estimation of the Doppler shift frequency in high-frame rate blood flow imaging with fewer pulse transmission. When an ultrasound pulse is transmitted to a large number of scatterers, the echoes from the scatterers overlap and interfere with one another. This interference causes a fluctuation in the phase of the received echo signal, and the estimation of the Doppler shift frequency will be disturbed by this phase fluctuation. A new technique proposed here makes it possible to eliminate the disturbed phase information by the interference, and the remaining information is applied to the estimation of the Doppler shift frequency. The instantaneous frequency of the echo signal can be used as an index of the influence of interference. To test this technique, radio-frequency echo signal from the carotid artery is used for the simulation, and the error of the estimated Doppler shift frequency is evaluated for several cases. The result shows that better performance can be achieved using this technique in the case of a limited number of pulse transmissions.

2:20

1pBB5. Ultrasonic technique spectram tomography. Kazuto Kobayashi (Honda Electron. Co., Ltd., 20 Oyamazuka Oiwa-cho Toyohashi, Aichi, Japan) and Sigeo Otuki (Tokyo Inst. of Technol., Kanagawa, Japan)

This is a study about a technique for distinguishing a speckle pattern by a fine structure of a size of about wavelength of the observation waves from a large structure image. Speckle pattern is the one of the major factors that disturbs the observation of fine structure. It is generated by interference caused from spacious frequency factor and not by existence of the object. This echo image forming apparatus comprises: a wave transmitter for transmitting an observation wave pulse of a wide band; a wave receiver for receiving an echo of the observation wave pulse transmitted to generate a reception signal; a filter for extracting signal components of a plurality of different frequencies predetermined from the reception signal obtained by the wave receiver; and a display device for imaging separately, respectively, the signal components of each frequency extracted by the filter and displaying images corresponding to resultant frequencies mutually relating to each other. The signal components of the plurality of frequencies predetermined are extracted from the reception signal of the wide band by use of the filter. The speckle patterns by the fine structure of about wavelength in correspondence with the corresponding frequency emerge respectively in each extracted frequency component.

2:35

1pBB6. Spectral characteristics of the sound generated by ultrasound imaging systems in the human body. Mostafa Fatemi, Paul L. Ogburn, and James F. Greenleaf (Mayo Foundation, Rochester, MN 55905)

Medical ultrasound scanners use high-energy pulses to probe the human body. It can be shown that the radiation force resulting from the impact of such pulses to the object can vibrate the object, producing a localized high-intensity noise in the audible range. When the scanner is used in a maternal examination, and the beam is directed toward the fetal head, this noise can disturb the fetus and cause it to move vigorously [J. Ultrasound Med. **20**, 883–889 (2001)]. Here, a theoretical model for the audio noise generated by ultrasound scanners is presented. This model describes the temporal and spectral characteristics of the noise. It has been shown that the noise has rich frequency components at the pulse repetition frequency and its harmonics. Experiments have been conducted in a water tank to measure the noise generated by a clinical ultrasound scanner. Results indicate that the noise pressure level at the focal point of the ultrasound beam on a reflective object can reach up to 120 dB relative to 20 μP. This result is in agreement with the theoretical predictions. The noise was also measured in a human uterus, indicating intensities in the range of 90–100 dB.

2:50

1pBB7. Influence of airway wall properties on respiratory system response characteristics. Ahmed Al-Jumaily and Younis Al-Fakhri (Diagnostics and Control Res. Ctr., Auckland Univ. of Technol., Auckland, New Zealand)

Changes of airway wall physical properties are associated with inflammation and swelling of the airway passages. This contributes to airway obstruction and increased work of breathing. A model is developed to study the effect of changes in the wall properties on the dynamic characteristics of the respiratory system. Both symmetric and asymmetric structures are considered. By developing some insight into the correlation between the respiratory system structure and its acoustical response, groundwork is set for the analytical interpretation of high frequency response data in clinical and laboratory tests. The response of the lung structure indicates that when the airway wall elastance and inertance are varied, as the case during asthma attack (airway walls get swollen results in airway constriction and tissue properties changing), the overall normalized input impedance frequency spectrum measured at the throat gives a reasonable signature for identifying such abnormality. The identification procedure is based on comparing the frequency spectrums of healthy and unhealthy lungs. The branching structure of the airway walls and its dynamical properties appear to significantly influence the lung response.

3:05

1pBB8. Computer modeling of swallow using acoustic measures. T. Neil McKaig (Univ. of Stuttgart, Azenbergstr. 12, D-70174 Stuttgart, Germany)

Although cervical level auscultation of the human swallow has long been discussed as a potential adjunct to the clinical dysphagia examination, there has been little research to quantify the acoustic parameters of the procedure. After initially examining the relative timing of these events in a normal population of +250 subjects, algorithms capable of generating computer models based on relative timing of acoustic data were developed. Procedure: In addition to physiological profile and medical/social history, a sample swallow is recorded and 6 raw measurements data points are identified: initiation of deglutition apnea (dA), leading and trailing edges of both the initial and secondary bursts as well as the terminal acoustic event. From this an additional seven measurements are derived: duration of dA, relative midpoint of dA, durations and midpoints of initial and secondary bursts, the interburst duration, and the relative midpoint of the bolus transport signal (BTS). Conclusion: It is possible for the computer to generate an optimal swallow profile for an individual and through statistical analyses predict the range, median and relative timing for each event reporting profile and the degree of confidence based on the subset n. [Equipment for capturing data and computer analysis provided by Welsh/Allyn Corporation.]

Session 1pMU

Musical Acoustics and Education in Acoustics: Musical Technology

Thomas D. Rossing, Chair
Physics Department, Northern Illinois University, DeKalb, Illinois 60115

Invited Papers

1:00

1pMU1. Music technology and how it is being taught. Thomas D. Rossing (Phys. Dept., Northern Illinois Univ., DeKalb, IL 60115)

Music technology addresses techniques for creating, recording, and reproducing sound, constructing and repairing musical instruments, developing new and improved instruments, creating new sounds with computers, and scientific research to support these endeavors. Although rather well-defined music technology programs exist in other countries of the world, in the United States these techniques are generally taught in several different disciplines, and music technology is an interdisciplinary field of study. At our university, for example, students generally major in either physics or music but take courses in engineering and computer science as well. Attractive opportunities await students who gain a good background in musical acoustics and music.

1:20

1pMU2. Teaching musical technology at the University of Edinburgh. D. Murray Campbell (Dept. of Phys. and Astron., Univ. of Edinburgh, James Clerk Maxwell Bldg., Edinburgh EH9 3JZ, U.K., D.M.Campbell@ed.ac.uk)

The teaching of musical technology at the University of Edinburgh dates from the appointment of John Donaldson as Professor of Music in 1845. One of the courses which he offered was on ''The Structure, Compass and Properties of Musical Instruments,'' the stated objective being ''to discover the true principles on which musical instruments ought to be constructed, and which may lead, and have led, to the invention of new ones.'' To this end he created a music classroom well equipped with acoustical apparatus and demonstration instruments. Out of Donaldson's pioneering work developed a long-standing cooperation between the Faculty of Music and the Department of Natural Philosophy (later Physics and Astronomy) at Edinburgh. Currently two Honors Degree programs with a strong emphasis on musical technology are offered. One is a Bachelor of Music Degree in Music Technology; the other is a Bachelor of Science Degree in Physics with Music. Both programs include courses in introductory musical acoustics, acoustics of musical instruments, analysis and synthesis of sound, and practical music technology. The B.Sc. requires an advanced school qualification in physics and mathematics, but no music; the B.Mus. requires an advanced school music qualification, but less advanced scientific background.

1:40

1pMU3. The music technology program at York. David M. Howard (Dept. of Electron., Univ. of York, Heslington, York YO10 5DD, UK)

For the last 8 years, undergraduate degrees (fully accredited by the Institution of Electrical Engineers) in Electrical Engineering with Music Technology Systems have been available in the Department of Electronics at the University of York (UK). These courses maintain a thread of music technology related material alongside the core electronics material throughout the 3 (BEng) or 4 (MEng) years of the program which includes acoustics and psychoacoustics, digital computer music, recording studio techniques, music technology ECAD (MATLAB) exercise, and advanced music synthesis. Students have to participate in a concert (as a group exercise), and in recording exercises (group and individual). Entry requirements are: three grade B passes at A-level including Mathematics and a science subject (usually Physics) as well as musical performance skills [Grade VII from the Associated Board of the Royal School of Music (ABRSM) or equivalent] and some music theory background (ABRSM Grade V Theory or equivalent). Employers have noted that graduates from these Music Technology programs have an added range of creative skills beyond music and audio that they can apply to other areas of engineering design, such as printed circuit board layout and user interface design.

2:00–2:15 Break

2:15

1pMU4. Interactive web-based software for musical technology. Daniel Boye (Phys. Dept., Davidson College, P.O. Box 7133, Davidson, NC 28036-7133)

A general course covering digital music basics, electronic synthesis of musical instruments, room acoustics, CD and DVD players, and compressed data formats has been developed to draw upon the penchant of college students for music and computer technology. Nonscience majors populate this course at Davidson College to satisfy an elective or a core requirement in the traditional liberal arts curriculum. Students directly apply material covered in the course to daily experiences and to musical performances they are required to attend. This presentation will briefly describe the course structure and focus on several interactive web-based sound exercises and demonstrations that examine acoustical and psychoacoustical phenomena. Interactivity is achieved through the use of Java-based Physlets (developed at Davidson), JavaScript, and streaming of audio clips. Currently available time and frequency domain visualization plugins for popular media players (Windows Media Player, WinAmp), while imaginative and colorful, do not lend themselves to scientific measurement. New measurement-friendly visualization plugins are being developed at Davidson. These plugins will be described and made available for download.

2:35

1pMU5. Music Technology at Northeastern University. Anthony De Ritis (Northeastern Univ., 351 Ryder Hall, Boston, MA 02115)

Inititated in 1998 by Professors Dennis Miller and Anthony De Ritis, the Music Technology program at Northeastern University requires all students to take private lessons in music composition. During the first two years, emphasis is on acoustic instruments only. Acoustics and psychoacoustics play an important role in this process. Music Technology students are required to take embedded learning modlues, or ELMOs, introducing concepts in the physics and perception of sound in a hands-on laboratory environment. We believe ELMOs shorten the learning curve students require for the acquisiton of advanced orchestration skills and serves to heighten their awareness in sound design. Core courses include Interactive Real-Time Performance (using the MAX programming language), Composition for Electronic Insturments, Advanced Music Systems (including Kyma and C-Sound), and Sound Design, which includes the production of an entire radio drama for its final exam. At the end of their sophomore year, students have the option to apply for the Multimedia Studies Dual Major, grouping majors in music technology, animation, graphic design and digital imaging in multimedia-based projects. The Music Technology program is a leader in Education Technology at Northeastern, and was recently awarded a $25 000 grant to develop an internet-based course called Computer Literacy for Musicians.

2:55

1pMU6. Evolution of teaching and research at CCRMA. Chris Chafe (CCRMA/Music, Stanford Univ., Stanford, CA 94305)

Three decades of research at the Center for Computer Research in Music and Acoustics have been accompanied with an evolving program for teaching music science and technology. As new understanding and new tools advance the research, a growing set of methods and skills need to be covered. The heavy emphasis on ''project-based'' teaching at CCRMA crosses several subdisciplines and serves students with a wide range of backgrounds and levels of experience. One thing has been constant throughout the program's existence: surprises happen in teaching. The talk will conclude with a description of ingredients that help encourage serendipitous results at the lab.

3:15

1pMU7. Soundtrack: Focus on electronics. Thomas L. Rhea (Berklee College of Music, 1140 Boylston St., Boston, MA 02215)

A retrospective-perspective on the nexus where music composition, electronic technology, and music technology converge to facilitate production of underscoring, foley, and sound effects for motion pictures. Clips excerpted from the speaker's film collection will be presented to illustrate: (1) the arrival of film sound, with a Vitaphone demonstration; (2) music editing, with illustrations of click track, and punches and streamers synchronization techniques; (3) production and layering of modern sound FX and foley; (4)

musical examples of ''Mickey Mousing,'' and source music vs underscore distinctions; (5) demonstrations and film clips featuring electronic musical instruments used in the film score, e.g., theremin and Hammond Novachord; and (6) assembly of modern soundtrack elements to present a sonic ''hyper-reality.'' The thesis that early sound film workers anticipated many of the magnetic audio tape techniques developed by pioneers in the early production of electronic music will also be explored. [Support in the form of stipend, travel, housing, and technical personnel support provided during a Residency in 2000–2001 at the Institute of Electronic Arts (IEA) at Alfred University, NY. Film clips presented on DVDs mastered by Ms. Lara Odell.]

3:35

1pMU8. Music Technology Program at McGill University. Philippe Depalle (Dept. of Theory, Faculty of Music, McGill Univ., 555 Sherbrooke St. W., Montreal, QC H3A 1E3, Canada, depalle@music.mcgill.ca)

The Music Technology program at McGill University encourages interaction between musical creation, technology, and research. It inherits from a long tradition of music and science combination, which was initiated in the early 1960s at Istvan Anhalt's instigation and thanks to Hugh Le Caine's contributions. The pedagogical goal of the program is to provide music students with an environment for professional-level music training and with an intensive focus on programming of advanced music technologies. The undergraduate level has been recently overhauled to reflect the current state of the field and the growing importance of multimedia. This led to a substantial increase of the number of students which will boost next year the population at the graduate level. This factor, in conjunction with the doubling of the academic staff and the creation of a new research center called Centre for Interdisciplinary Research in Music, Media, and Technology (CIRMMT) will reinforce the research activity in the field. CIRMMT also combines several research efforts, which were previously distributed around McGill University (in psychoacoustics, acoustics, cognitive sciences, and electrical engineering applied to music) into a creative interdisciplinary entity that will be housed within the new Faculty of Music Building.

3:55

1pMU9. The teaching of musical technology at London Guildhall University. Lewis Jones and Allan Seago (London Guildhall Univ., 1 Commercial Rd., London E1 1LA, UK, ljones@lgu.ac.uk)

London Guildhall University has recently fundamentally reformed its undergraduate teaching of musical technologies. There are now two broad streams, one concentrating on the study, making, design, and restoration of musical instruments, and one on audio systems. Central to both is teaching in acoustics and a range of course units which explore the relationship between music and technology in its historical, contemporary, and ethnographic aspects. New MMus and MSc courses in ''Music and Technology'' will provide research training and opportunities for advanced project work. The rationale for these developments is accounted for, and the approach to the teaching of musical acoustics to music technologists and instrument technicians is described. A number of Java-based programs and Quicktime animations have been developed. The animations serve simply to demonstrate such phenomena as simple harmonic motion and standing waves. The Java applet which has been developed by Allan Seago is intended to demonstrate resonance, and the factors which affect it. Unlike the animations, the software is interactive, and enables the student to vary such aspects of the system as mass, spring constant, and damping, and to observe the effect.

Contributed Papers

4:15

1pMU10. Music synthesis program at Berklee College of Music. Thomas L. Rhea (Berklee College of Music, 1140 Boyston St., Boston, MA 02215)

A description is given of the Music Synthesis Department of the Music Technology Division of Berklee College of Music in Boston, MA. A ''virtual tour'' of facilities will be provided, as well as a brief exposition of the teaching and support infrastructures. The session will comprise examples of Music Synthesis Department (undergraduate) student compositions, soundscapes, web page designs, and multi-media productions, and research projects in acoustics and psychoacoustics completed as final projects required for graduation. Questions from any educators interested in logistics for this type of program will be answered.

4:30

1pMU11. Acoustics, Audio, and Music Technology Education at the University of Illinois at Urbana-Champaign. James W. Beauchamp (School of Music and Dept. of Elec. Eng., Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801, j-beauch@uiuc.edu), Heinrich Taube, Sever Tipei, Scott A. Wyatt (School of Music, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801), Lippold Haken, and Mark Hasegawa-Johnson (Dept. of Elec. Eng., Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

Several courses combine to form an unofficial but nevertheless effective undergraduate and early graduate music technology program at the University of Illinois at Urbana–Champaign. *Musical Acoustics* is taught within the School of Music and includes topics such as resonance behavior of strings and pipes; perception of loudness, pitch, and timbre; and acous-

tics of musical instruments. Several courses involving computers and music are taught in the School of Music. These courses discuss the fundamentals of computer music, and extensive computer/recording studios provide students with hands-on opportunities to conduct experiments or to produce audio demonstrations or music compositions. The courses differ in terms of the user interfaces, compositional algorithms, and synthesis models employed. There is also a seminar in analysis and synthesis of sounds. In the Department of Electrical and Computer Engineering, besides conventional digital signal processing, there are several special courses that offer training for a career in music technology. *Fundamentals of Engineering Acoustics* supplies a rigorous introduction to wave propagation and acoustical systems. *Electronic Music Synthesis* describes algorithms for music sound synthesis and representation and includes a laboratory component. *Topics in Audio Engineering* covers psychoacoustics, signal processing, audio coding and recognition, room acoustics, and tranducers.

MONDAY AFTERNOON, 3 DECEMBER 2001 ROOM 222, 1:00 TO 4:30 P.M.

Session 1pPA

Physical Acoustics: Applications of Acoustic Resonance

Joseph R. Gladden, Chair
Department of Physics, Pennsylvania State University, 104 Davey Laboratory, State College, Pennsylvania 16802

Invited Papers

1:00

1pPA1. Acoustic resonators for measuring the thermodynamic and transport properties of industrial gases and for calibrating thermometers. Michael R. Moldover (Process Measurements Div., Natl. Inst. of Standards and Technol., Gaithersburg, MD 20899)

At the NIST, high Q cylindrical acoustic resonators are routinely used to measure the speed of sound in difficult-to-handle process gases with uncertainties of 0.01% or less. Model intermolecular potentials are fitted to the acoustic data to obtain the gas density and heat capacities with uncertainties of 0.1% as well as estimates of the viscosity and thermal conductivity. We have developed novel, low-Q, acoustic resonators to directly measure the viscosity and thermal conductivity. Spherical resonators are used to measure the speed of sound in gases with sub-part-per-million uncertainties. An argon-filled spherical resonator was used to redetermine the universal gas constant R with a fractional standard uncertainty of 1.7×10^{-6}. The same resonator was used to measure imperfections of the internationally accepted temperature scale (ITS90) in the range 217–303 K with uncertainties of 0.6 mK; this work is now being extended to 800 K. In effect, very accurate speed-of-sound measurements will be used to calibrate other thermometers.

1:30

1pPA2. Application of acoustic resonators in photoacoustic trace gas detection and metrology. Peter Hess (Inst. of Physical Chemistry, Univ. of Heidelberg, 69120 Heidelberg, Germany)

The application of different types of acoustic resonators such as pipes, cylinders, and spheres in photoacoustics will be considered. This includes a discussion of the fundamental properties of these resonant cavities. Modulated and pulsed laser excitation of acoustic modes will be discussed. The theoretical and practical aspects of high-Q and low-Q resonators and their integration into complete photoacoustic detection systems for trace gas monitoring and metrology will be covered. The main characteristics of the available laser sources and the performance of the photoacoustic resonators, such as signal amplification will be discussed including setup properties and noise features. Recent results will be presented for a state-of-the-art dual-resonator differential cell suitable for sensitive trace gas detection with low electronic and acoustic noise. With this resonant cell and a near-infrared DFB diode laser radiating at 1.53 μm (42 mW) polar ammonia molecules have been detected with a sensitivity of 200 ppbv under flow conditions to reduce adsorption effects. By excitation of fundamental vibrations in methane using a pulsed optical parametric oscillator (OPO), with 60 mW output power, a sensitivity of 1.2 ppbv has been achieved employing this resonant cell. This setup allows accurate detection of the greenhouse gas methane which has a concentration of about 1.7 ppmv in ambient air.

2:00

1pPA3. A review of elastic constant measurements with resonant ultrasound spectroscopy. Jin H. So, J. R. Gladden, and J. D. Maynard (The Pennsylvania State Univ., University Park, PA 16802)

Measurements of elastic constants in new materials are crucial because the elastic constants, being directly related to atomic and electronic configurations, carry important information about the mechanical and electronic properties and phase transitions. A conventional method for measuring elastic constants, the pulse-echo technique, suffers from a number of problems, such as transducer-sample coupling, parallelism of sample faces, the necessity of remounting transducers, etc. Such problems are avoided in resonant ultrasound spectroscopy (RUS), a relatively new method which can determine all elastic constants in a single measurement. Currently a number of research groups, e.g., UCLA, Los Alamos, Colorado State, University of Wisconsin, University of Nevada, NIST, and laboratories in Europe and Asia are using the RUS technique to study a wide variety of systems, and these will be reviewed in this talk. In our own research, the RUS technique has been extended for studies of very small samples (less than 100 μg) and for measuring the elastic constants in thin film (200 nm) samples. Studies include thin films of magnetic materials involving colossal magnetoresistance (candidates for high density digital data storage media) and films of carbon nanotubes, which have unique mechanical properties. [Research supported by ONR.]

2:30

1pPA4. The use of acoustic resonance in particle manipulation and separation. Todd L. Brooks and Robert E. Apfel (Dept. of Mech. Eng., Yale Univ., New Haven, CT 06520-8286)

Particles in an acoustic standing wave experience a force directed towards pressure nodes or antinodes depending upon particle and host fluid properties. The observation of cork dust banding in air was first reported in 1866 by Kundt, but only recently has there been a renewed interest in utilizing such acoustic forces as a means for practical separation processes in liquids. Applications in biotechnology include two-phase partitioning, agglutination, electro-acoustic cell fusion, and cell retention systems. This talk describes a continuous flow stream with a perpendicularly oriented standing wave in a half or quarter wavelength resonator. This allows a continuous separation process in which the output flow can be split into several streams of different particle concentration, type, or size. Since the acoustic force depends on a particle's compressibility (as well as size and density contrast), there exists a unique opportunity to separate particles based on this property, not currently available in other separation methods. Competing effects, such as thermal convection, acoustic cavitation, and streaming, reduce separation efficiency, and these must be understood and controlled. A survey of applications will be presented along with recent experimental work and numerical simulations done in the Yale Acoustics Laboratory. [Work supported by NASA through Grant No. NAG8-1351.]

3:00–3:15 Break

Contributed Papers

3:15

1pPA5. Acoustic resonators as air quality sensors. Ralph T. Muehleisen (Dept. of Civil, Environ., and Architectural Eng., Univ. of Colorado, Boulder, CO 80309)

The past decade has seen a dramatic rise in concern for the quality of air that we breathe in our homes and workplace. Inexpensive sensors to measure changes of the composition of air are required. One can measure changes in the speed of sound of air by measuring the change in the resonance frequency of an open air acoustic resonator. If the air is treated as an ideal gas, changes in the speed of sound are directly related to changes in the gas composition. This presentation will discuss common indoor contaminants and the concentrations at which they contribute to poor indoor air quality (IAQ) and occupational safety. It is found that acoustic resonant sensors are not sensitive enough for use as IAQ sensors but are sensitive enough to be used as occupational safety monitors. [Work supported by ONR.]

3:30

1pPA6. Use of radiation pressure to move small particles in an air flow-through resonator. Ralph S. Budwig, Joseph G. Frankel, Michael J. Anderson, and K. Scott Line (Dept. of Mech. Eng., Univ. of Idaho, Moscow, ID 83844-0902)

It has been demonstrated many times in the past that particles suspended in a fluid medium can be manipulated with radiation pressure. Resonant cavities are often used to achieve large acoustic amplitudes. While most investigations have used water as the fluid medium, few have reported using a gas such as air. Little is known about the effect of openings, required for mean fluid flow through the cavity, on acoustic resonant amplification. Finally, few investigations report actual sound pressures that can be achieved, and correlate them to the movement of the particles. We present experimental measurements that quantify the ability of a flow-through resonator to move small water particles in air as the flow velocity and sound pressure levels are changed. The small water particles were generated with an ultrasonic humidifier. A commercial electrostatic transducer was employed to excite a one-wavelength longitudinal resonance of the cavity at a frequency near 60 kHz. Preliminary results show that sound pressure levels in excess of 150 dB re: 20 μPa can be attained, and significant particle movement is observed at flow velocities up to 11 cm/s. Observed particle movements will be compared to simple theoretical predictions based upon measured sound pressure levels and flow velocities.

3:45

1pPA7. Resonant sonic spectroscopic studies of aerogel solids. Edward Tucholski, Andres Larraza (Naval Postgrad. School, Phys. Dept., Monterey, CA 93943, larraza@physics.nps.mil), and Michael Droege (Ocellus Technologies, Livermore, CA 94550-9291)

Measurements similar to those performed using resonant ultrasound spectroscopy (RUS) were carried out in the audible frequency range for isotropic silica aerogels in air. This solid material has attracted significant attention due to its interesting physical properties. It has a density only 100 times greater than air, an optical index of refraction nearly equal to air (1.0–1.05), and a thermal conductivity of ~0.017 W/mK (~R10/in.). The acoustic measurements, conducted on a series of aerogel spheres ranging in diameter from 1.1 to 4.6 cm, demonstrated the two elastic moduli are four orders of magnitude less than that of dense silica and correspond to a longitudinal wave speed one-third that in air. The results are compared to pulse-receiver sound speed measurements. Sample spherical imperfections are demonstrated.

4:00

1pPA8. Radiation-pressure induced frequency shift of liquid capillary bridge transverse modes: A string approximation. Philip L. Marston and David B. Thiessen (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

Liquid capillary bridges in air have been stabilized by placing them at the velocity anti-node of an ultrasonic standing wave. This radiation-pressure induced stabilization depends on an acoustic parameter q that is proportional to the square of the ultrasonic pressure amplitude [M. J. Marr-Lyon *et al.*, Phys. Rev. Lett. **86**, 2293–2296 (2001)]. An approximation developed here suggests that an independent measure of q could be obtained by measuring the increase in the natural frequency of the lowest transverse bridge mode. That mode is known as the (1,1) mode. Even for an inviscid bridge, the exact (1,1) mode description is complicated. This complication is avoided by showing that the natural frequency for long round bridges is approximately that of a string having an effective tension CT/2 where C is the bridge circumference and T the surface tension. The additional restoring force (associated with the acoustic radiation pressure) for a bridge at a velocity anti-node is modeled as a uniform elastic support. When q is small, the frequency shift is predicted to be proportional to q. The calculation should also apply to the use of vibrating strings to measure acoustic standing wave amplitudes. [Work supported by NASA.]

4:15

1pPA9. Acoustic resonance determination of the rigidity modulus. Kenneth G. Foote (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Acoustic resonance methods are acknowledged to be particularly powerful in the determination of elastic properties of materials. Three measurement techniques for determining the rigidity modulus are described.

These involve comparison of the echo spectrum, energy, or waveform with the respective theoretically computed quantity, where the rigidity modulus, or shear-wave sound speed, is a free parameter. The value of this parameter is adjusted to achieve the best fit with measurement. The second technique is illustrated to specify the rigidity modulus of electrical-grade copper. An echo sounder operating at 120 kHz, calibrated by reference to the echo from a 30.05-mm-diam copper sphere, is used to excite a resonance in a 60.07-mm-diam sphere of the same material. The rigidity modulus is found to be 46.86 ± 0.06 GPa and shear-wave sound speed, 2288 ± 2 m/s.

Session 1pSA

Structural Acoustics and Vibration: Energy-Based Methods

Rudolph Martinez, Chair
Cambridge Acoustical Associates/Anteon Corporation, 84 Sherman Street, Cambridge, Massachusetts 02140

Contributed Papers

2:00

1pSA1. Predicting variability in SEA. Richard H. Lyon (RH Lyon Corp, 691 Concord Ave., Cambridge, MA 02138)

The variability that exists between an SEA estimate of the response of an ''average'' system and the actual response of a particular system has been estimated in various papers and books by the author. This estimate is in the form of a central limit theorem in which the variance, normalized by the square of the mean, is reduced by a factor that is the number of independent interacting mode pairs. A review of that result indicates that this parameter is overestimated, resulting in an underestimate of the actual variance. A new estimate of the variance is derived, which appears to have the correct asymptotic behavior under conditions of both low and high modal overlap. [Research supported by Cambridge Collaborative, Inc.]

2:15

1pSA2. Radiation from nonhomogeneous structures using SEA. Jerome E. Manning (Cambridge Collaborative, Inc., 689 Concord Ave., Cambridge, MA 02138)

Prediction of the acoustic radiation from vibrating structures at high frequencies can be accomplished using statistical energy analysis models. However, the basic formulations of radiation efficiency used in these models are often based on assumptions regarding boundary conditions and structural homogeneity that are not valid. Recent work carried out for nonhomogeneous structures with realistic boundary conditions is presented. An impedance approach is used that allows the radiation to be defined in terms of the forces acting on the structure from boundaries and added components. Examples are used to demonstrate the changes in radiation efficiency resulting from realistic boundaries, frame stiffeners, and added components.

2:30

1pSA3. Experimental statistical energy analysis in the time domain. Joseph Gregory, Richard Keltie, and Hubert Hall (Dept. of Mech. & Aerosp. Eng., North Carolina State Univ., Raleigh, NC 27695-7910, keltie@eos.ncsu.edu)

A technique for identifying an experimentally based power flow model from transient measurements is described. The technique was validated using a range of test simulations and then with true physical tests conducted on a structure comprised of coupled thin plates. The identified power flow model was used to successfully update and improve an existing computational statistical energy analysis (SEA) model. It was found that utilization of time domain data allowed for an overdetermined power balance providing favorable numerical conditions for the identification. Useful insight into system order/dimension was obtained during the identification process. It was observed that a full matrix of measured inputs and outputs was not required. However, inclusion of the extra measurements was seen to further increase overdetermination, yielding an even better least squares estimate of model parameters. It was found that the identification procedure indicated true parameters that were easily distinguished from those associated with noise in the data and, hence, seemed well suited for this SEA application. Results indicate that the methodology has the potential to enhance standard experimental SEA procedures.

2:45

1pSA4. Provided the modal overlap parameters exceed a threshold, statistical energy analysis (SEA) is not challenged by energy analysis (EA). G. Maidanik and K. J. Becker (NSWCCD, David Taylor Model Basin, 9500 MacArthur Blvd., West Bethesda, MD 20817-5700)

A structural complex is modeled by an externally driven master oscillator that is coupled to a set of satellite oscillators. In this model the satellite oscillators are not coupled to one another and the resonance frequencies of the satellite oscillators centrally span that of the master oscillator. The response of the model is cast in terms of the modal coupling strength $\zeta(y)$. This quantity is the ratio of the energy typically stored in a satellite oscillator to the energy stored in the master oscillator, (y) is the normalized frequency (ω/ω_0) and (ω_0) is the resonance frequency of the master oscillator. The modal coupling strength $\zeta(y)$ determined via the statistical energy analysis (SEA) is invariably less than unity; $\zeta(y)<1$. This quantity determined via the energy analysis (EA) is not necessarily so restricted. Indeed, the (EA) $\zeta(y)$ can be maintained below unity only if a typical modal overlap parameter (b_s) of the satellite oscillators exceeds a threshold (b_t); $b_t<b_s$. This threshold is required if the validation of SEA is not to be challenged by EA. Cited is a reasonable complex with strong couplings; e.g., satellite oscillators as sprung-masses, for which (b_t) exceeds unity. Therefore, to validate the SEA, modal overlap in this set must prevail.

3:00

1pSA5. Application of the mode isolation algorithm for experimental modal analysis to a system with close natural frequencies. Jerry H. Ginsberg and Bassem R. Zaki (G. W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

Experimental modal analysis techniques in current use have difficulty with modes whose natural frequencies are very close or whose drive point mobility is very low. How well these issues are addressed by the Mode Isolation Algorithm (MIA), see M. V. Drexel and J. H. Ginsberg, ''Modal parameter isolation using state space mode isolation,'' in Proceedings of the 19th International Modal Analysis Conference, Orlando, Florida, February 5–8, 2001, is the topic of this presentation. MIA is an iterative procedure that employs frequency domain response data. The application

of MIA to a system consisting of a beam with three suspended spring-mass-dashpot subsystems is described. Judicious choice of the system parameters leads to two modes having close natural frequencies, and two other modes in which impulse excitation at the free end of the beam results essentially in vibration of the subsystems away from the end. Synthetic impulse response data for the analysis is contaminating with white noise. It is shown that MIA accurately identifies all natural frequencies and modal damping ratios if the frequency resolution of the response data is sufficiently fine. Furthermore, the two normal mode shapes associated with the proximate natural frequencies are correctly identified, but the two with low drive point mobility are not. [Work supported by the National Science Foundation, Grant No. CMS-0098539.]

3:15

1pSA6. The formation of energy pass bands from periodically spaced point masses on a ribbed cylindrical shell. Martin H. Marcus and Brian H. Houston (Naval Res. Lab., 4555 Overlook Ave., S.W., Washington, DC 20375, marcus@astro.nrl.navy.mil)

An infinite cylindrical shell has bending energy that is able to propagate down the shell at all frequencies with little attenuation other than from material damping and the surrounding medium. In contrast, making the shell finite causes bending resonances to occur only at discrete frequencies. The addition of periodically spaced circumferential stiffeners causes the bending wave dispersion curves to be seen at additional axial modes, which is referred to as Brillouin folding. The stiffeners also cause energy stop bands to occur when the load is a single Fourier mode. In this presentation, periodically spaced point masses in the circumferential direction are added to a ribbed finite cylindrical shell and analyzed via the finite element method. It is seen that Brillouin folding occurs circumferentially. Also, energy pass bands are created in a certain pattern. Analysis of the axisymmetric version of this shell combined with information on the type of aperiodicity yields a prediction of the energy pass band behavior. This provides a qualitative approach toward understanding the motion of nonaxisymmetric shell structures. [Work supported by ONR.]

3:30–3:45 Break

3:45

1pSA7. Including acoustics in the RESOUND hybrid approach to the mid-frequency problem. Philip J. Shorter and Bryce K. Gardner (Vibro-Acoust. Sci., 12555 High Bluff Dr., Ste. 310, San Diego, CA 92130, pj.shorter@vasci.com)

The prediction of the response of a complex structural acoustic system across a broad frequency range presents a number of challenges to an analyst. It is quite common to find that (i) the uncertainty associated with the local modal properties of a subsystem, (ii) the wave number content of the local modes, and (iii) the local modal density, can all vary significantly across the various subsystems of a system. This mismatch in the local statistical and dynamic properties of a system is often referred to as the mid-frequency problem. One approach to the mid-frequency problem is to construct a hybrid model that combines deterministic and statistical representations of the system dynamics. Previously, this approach has been applied to *in-vacuo* structural problems. In this presentation we discuss the extension of the hybrid formulation to encompass enclosed acoustic cavities. The various mechanisms by which energy flows between the deterministic and statistical parts of the structural–acoustic model are discussed, and numerical examples are presented that illustrate radiation into a large (statistical) acoustic cavity and a small (deterministic) acoustic cavity.

4:00

1pSA8. Experimental evaluation of coupling loss factors of thin plates. Ryszard Panuszka, Jerzy Wiciak, and Marek Iwaniec (Staszic Univ. of Krakow, 30 Al. Mickiewicza, Krakow Pl 30-059, Poland, panuszka@purdue.edu)

Using statistical energy analysis method to gain knowledge of coupling loss factors (CLFs) of structural elements of complex mechanical systems is fundamental, therefore the evaluation of CLFs for different constructional connections of thin rectangular plates merits studying. CLFs for different constructional perpendicular connections of thin rectangular plates were investigated. Three types of couplings were studied: line junction, stiff bridges, and point connections. Investigations were realized with method of energy stored at anechoic chamber. Experimental test included two kinds of connections of tested perpendicular plates: welded line junctions and point junctions. Several combinations of thickness ratio on CLFs values were tested for welded plates. Also, the distribution points were studied to show the impact on CLFs. There is evidence of decreasing of CLFs for larger ratios of thickness especially for higher frequency bands. Acoustic power flow across riveted junctions shows up to 20 dB lower transmission than similar spot welded junctions. In the examined screwed junctions demonstrate lower than average CLFs activity than riveted and higher than point welded.

4:15

1pSA9. Vibrations models for offshore structures reinforcing beams. Alexandre Sardá, Arcanjo Lenzi (LVA, Dept. de Engenharia Mecânica, Univ. Federal de Santa Catarina, Caixa Postal 476, Florianópolis, SC, Brasil 88040-900), and Joseph Cuschieri (Florida Atlantic Univ., Dania Beach, FL 33004-3023)

Offshore structures used in oil prospection and production are mainly made of plates reinforced by beams. Vibrations produced by machines and equipment attached to their main structure propagate through such structural components generating usually high-noise levels at the accommodation areas. Power flow analyses through beam reinforced plates require proper models for L and T beams, in order to take into account the influences of their own modes formed in the web and flanges. This work presents a model for vibrations of a typical reinforced beam, made of 2 (or 3) plates welded in L (or T) shape, representing web and flange. Flexural and in-plane shear waves are considered in the plates. Results are compared to finite elements results indicating very good agreement and significantly reduced computational costs.

4:30

1pSA10. Vibration of membrane excited by coupled acoustic energy from the bulk substrate. Hang Guo and Amit Lal (SonicMEMS Lab, Dept. of Elec. and Comput. Eng., Univ. of Wisconsin—Madison, 1415 Engineering Dr., Madison, WI 53706)

Vibration of a membrane is widely used in MEMS devices, such as membrane pumps for microfluidics and rf membrane switches. Conventionally, the membrane is excited by using interdigitated electrodes on piezoelectric ZnO thin films on the membrane. The small piezoelectric coefficient of ZnO limits the acoustic intensity for actuator application and membrane vibrate at one fixed frequency since the distance between interdigitated electrodes specifies the wavelength. In this paper, a novel cost-effective method to excite the membrane is proposed. A PZT plate is bonded to the bulk substrate and used to excite the membrane. The fundamental of this method is that the acoustic energy in the bulk substrate is coupled to the membrane for its vibration through the connecting boundary between the membrane and bulk substrate. Finite element method is used to calculate resonant frequencies and coupled acoustic energy, while the laser interferometer is used to verify resonant frequencies and mode shapes obtained from FEM and measure the actual displacement of the membrane in the air. The results show that for a 2 mm$\times$2 mm$\times$2.232 μm membrane, more than 30 membrane vibrations are obtained within the frequency up to 3 MHz, and the biggest membrane displacement is 0.27 μm, which is two times bigger than the conventional method.

4:45

1pSA11. The electro-mechanical impedance method for damage identification in circular plates. Andrei N. Zagrai (Dept. of Mech. Eng., Univ. of South Carolina, Columbia, SC 29208, anzagrai@engr.sc.edu) and Victor Giurgiutiu (Dept. of Mech. Eng., Univ. of South Carolina, Columbia, SC 29208, victorg@sc.edu)

Among structural acoustics methods, the electromechanical (E/M) impedance method is a new method that is easy to implement on complex structures with minimum instrumentation and straightforward results. This new acoustic method benefits from the advent of commercially available low-cost piezoceramics that have opened new opportunities for structural identification. The structural dynamics is identified by obtaining the E/M impedance or admittance signatures of piezoelectric active sensors that are permanently attached to the structure. In the presented study, the E/M impedance method was used to identify the structural dynamics and the presence of damage in circular plates. An analytical expression, which accounts for both axial and flexural vibrations of a plate, was derived and validated through a set of experiments. The experimental results show good matching with the theory. Additional acoustic experiments were conducted to assess the presence of damage in circular plates for a particular set of boundary conditions. Some damage metrics were suggested to analyze the frequency spectrum obtained. The results show the changes in the E/M impedance spectrum due to damage presence. Future research will focus on the application of the E/M impedance to the acoustical diagnostics of aging aircraft structures.

MONDAY AFTERNOON, 3 DECEMBER 2001 ROOM 305, 1:15 TO 3:20 P.M.

Session 1pSPa

Signal Processing in Acoustics: Topical Meeting on Acoustic Time Reversal and Applications: Time Reversal in the Ocean Environment

Darrell R. Jackson, Chair

Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105-6698

Chair's Introduction—1:15

Invited Papers

1:20

1pSPa1. Applications of time reversal in ocean acoustics. W. A. Kuperman, W. S. Hodgkiss, Hee Chun Song (Marine Phys. Lab., Univ. of California, San Diego, La Jolla, CA 92093-0238), and Tuncay Akal (SACLANT Undersea Res. Ctr., 19138 La Spezia, Italy)

Electromagnetic, early ocean acoustic experimental and theoretical work as well as laboratory ultrasonics [Zeldovich *et al.*, *Principles of Phase Conjugation* (Springer-Verlag, Berlin, 1985); A. Parvelescu and C. S. Clay, Radio Electron. Eng. **29**, 223–228 (1965); D. R. Jackson and D. R. Dowling, J. Acoust. Soc. Am. **89**, 171–181 (1991); M. Fink, Sci. Am. Nov. 91–97 (1999)] suggest that time-reversal (or phase conjugation) methods provide opportunities for an assortment of applications in underwater acoustics. Recently, a series of ocean acoustic experiments have been carried out confirming the robustness and potential utility of time reversal mirrors in underwater acoustics [Kuperman *et al.*, J. Acoust. Soc. Am. **103**, 25–40 (1998); Akal *et al.*, 5th ECUA, Lyon, France, 2000]. This total sequence of research is briefly reviewed as an introduction to the research presented in this session. [Work supported by ONR.]

1:40

1pSPa2. Time-reversal in a waveguide. Philippe Roux and Mathias Fink (Laboratoire Ondes et Acoustique, ESPCI, 10 Rue Vauquelin, Paris 75005, France)

Experimental evidence of time-reversal has been performed at an ultrasonic scale (1.5–3.5 MHz) in a 70 long cm, 1 cm deep waveguide. This laboratory experiment allow us to get rid of noise or ocean fluctuations classically observed. Spatial and temporal time-focusing are studied and interpreted by the method of images in a waveguide or matched-field theory. For a wide band source array, an algorithm based on time-reversal is presented to achieve depth and range shifting without any knowledge of the waveguide properties.

2:00

1pSPa3. Modeling acoustic time reversal in random shallow water sound channels. David R. Dowling (Dept. of Mech. Eng., Univ. of Michigan, Ann Arbor, MI 48109-2121, drd@umich.edu)

The robust spatial and temporal focusing properties of time-reversing arrays (TRAs) are limited or lost when the environment is dynamic, acoustic absorption losses are prevalent, or noise contaminates the received signal(s). This talk presents theoretical and simulated performance results for TRAs deployed in shallow ocean sound channels containing a random linear superposition of internal waves or having a random rough bottom. In both cases, TRA retrofocusing is improved by random propagation and vertical

linear TRAs are predicted to have significant azimuthal directivity. Scaling laws for this azimuthal directivity that include the signal frequency, the characteristics of the sound channel, the properties of the randomness, and the assumption of weak azimuthal coupling are presented. These scaling laws are found to successfully collapse results for a wide parametric range of N by 2-D parabolic equation simulations of TRA retrofocusing in shallow ocean sound channels with random propagation. As expected, the TRA retrofocusing improvements derived from random propagation are typically degraded and are eventually lost when the environmental randomness is time varying. However, the time scales of the oceanic fluctuations appear to be long enough to allow successful TRA operations in shallow ocean waters. [Work sponsored by ONR.]

2:20

1pSPa4. Application of time reversal to reverberation reduction in the ocean. Seongil Kim (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238, seongil@mpl.ucsd.edu)

Reverberation from rough ocean boundaries causes serious problems in many active sonar applications. Numerous passive signal processing techniques have been developed to effectively extract target signals from reverberation noise. The spatial and temporal focusing capability of the time-reversal method provides a new approach to this problem by actively minimizing the reverberation. A time-reversal process realizes a focus at the probe source position and shadows the boundaries below and above the focus resulting in reduced reverberation. Recent experiments with 3.5-kHz sound showed the possibility of the time-reversal method to the detection problem by demonstrating the enhanced echo-to-reverberation ratio in shallow water. A time-reversal based process for reverberation cancellation from a specified range cell without the use of a probe source has also been developed. An eigen-decomposition of the reverberation data provides the transfer function between the TRM and scatters where the reverberation is generated [Prada *et al.*, J. Acoust. Soc. Am. **99**, 2067–2076 (1996); Lingevitch *et al.*, J. Acoust. Soc. Am. **109**, 2495 (2001)]. The backpropagated field from TRM with a weight function obtained by projection operations to the null space places an acoustic null at the corresponding boundary. This process results in reduced reverberation returning from the boundary. [Work supported by ONR.]

2:40

1pSPa5. Application of time reversal to underwater acoustic communications. Geoffrey Edelmann (Marine Physical Lab., Scripps Inst. of Oceanogr., 9500 Gilman Dr., La Jolla, CA 92093-0238)

Underwater acoustic communication systems must mitigate the intersymbol interference caused by the time-varying multipath dispersion. Time-reversal (TR), or phase-conjugation, focuses energy back to its origin despite the complexity of the propagation channel. The TR properties of dispersion reduction and the mitigation of the effects of channel fading make time-reversal a potentially attractive component of a coherent communications system. An overview of TR ocean acoustics concepts illustrated with data is presented as related to underwater acoustic communications. Experimental results of TR and control transmissions of binary phase shift keying (BPSK) and quadrature phase shift keying (QPSK) communication sequences are also presented together with measured bit error rates. These results will be discussed in relation to environmental fluctuations and their effects on the time-reversal focus. Extended duration (tens of minutes) time-reversal pulse transmissions will be examined as well as the time-evolving characteristics of the channel impulse response. [Work supported by ONR.]

3:00

1pSPa6. Passive phase conjugation for underwater communication. Daniel Rouseff, Warren L. J. Fox, and Darrell R. Jackson (Appl. Phys. Lab., College of Ocean and Fishery Sci., Univ. of Washington, Seattle, WA 98105, rouseff@apl.washington.edu)

Active phase conjugation requires an array capable of both transmitting and receiving. A field incident on the array can be refocused both in space and time at the location of the original source. To do acoustic communication, an additional step in the processing is introduced; prior to backpropagation, the measured probe field is first convolved with a data stream. The direction of communication is from the active array to the location of the original point source. By contrast, in passive phase conjugation [D. R. Jackson *et al.*, J. Acoust. Soc. Am. **108**, 2607 (2000)] the direction of communication is from the point source to the passive array. Further results from an experiment conducted in Puget Sound are presented. The effects of array curvature and truncation are discussed. [Work supported by ONR.]

Session 1pSPb

Signal Processing in Acoustics: Topical Meeting on Acoustic Time Reversal and Applications: Time Reversal in Random and Reverberating Media

George C. Papanicolaou, Chair
Department of Mathematics, Stanford University, Stanford, California 94305

Invited Papers

3:35

1pSPb1. Imaging and time reversal in random media. George C. Papanicolaou (Dept. of Mathematics, Stanford Univ., Stanford, CA 94305)

Time reversal in random media can result in refocusing on the source with a resolution that is better than the diffraction limit. This is because of the multipathing that occurs in the random medium, making the time reversal array appear to be wider than it is. What is not clear, however, is why this phenomenon is statistically stable. That is, why the refocused spot size does not fluctuate in the random medium. The reason for this statistical stability is that it is a time domain phenomenon and does not occur with narrow band signals. I will explain this effect analytically and will also show how it can be used for imaging in random media.

3:55

1pSPb2. Long distance time reversal and imaging in random media: Numerical simulations. Peter Blomgren, George Papanicolaou (Mathematics, Stanford Univ., Stanford, CA 94305, blomgren@math.stanford.edu), and Hongkai Zhao (Mathematics, Univ. of Calif., Irvine, Irvine, CA 92697-3875)

We have conducted a careful time reversal and imaging study for long distance wave propagation in random media in the underwater acoustics regime. In this talk we briefly describe the computational technology, which is based on a phase-screen propagation code for the paraxial wave equation. The advantages and limitations of the code are discussed. We show some time reversal results which demonstrate the super-resolution phenomenon and its statistical stability in the time domain. We also show results of imaging in random media using an array of active transducers.

4:15

1pSPb3. Communication in a diffuse random medium with time-reversal techniques. Arnaud Derode, Arnaud Tourin, Mickael Tanter, and Mathias Fink (LOA, 10 rue Vauquelin, 75005 Paris, France)

The application of time-reversal focusing to communication in a diffuse random medium is discussed, and new experimental results are presented. Binary information was transmitted by using a 24-element time-reversal array, with a 3.2 central frequency, either in water or in a multiple scattering sample. The media we studied consisted of random collections of parallel steel rods immersed in water. Experiments show that it is possible to take advantage of high-order multiple scattering in order to transmit five different binary sequences to five different points simultaneously, which would be impossible to achieve in a homogeneous medium with the same antenna, due to its poor resolution. The results are discussed in relation with the numbers of significant singular values of the propagation matrix.

4:35

1pSPb4. Time reversal through a solid–liquid interface and super-resolution. Chrysoula Tsogka (CNRS/LMA, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France) and George C. Papanicolaou (Dept. of Math., Stanford Univ., Stanford, CA 94305)

Time-reversal can be used to refocus optimally a signal back onto the source that emitted it. An active array antenna is used to record the wave as a function of time and then, the recorded field is time reversed and retransmitted. The time reversed wave backpropagates through the medium and refocuses approximately on the initial source position. From the experimental point of view, time-reversal devices are designed to work primarily in a fluid environment. However, time reversal refocusing can be realized also

in solids. To do so, the solid is surrounded by a fluid in which the active antenna is located. We propose a numerical formulation describing the time reversal phenomenon through a fluid–solid interface and we analyze with numerical simulations the refocusing resolution obtained in homogeneous media. We also show numerically that multipathing caused by random inhomogeneities improves the focusing of the back-propagated elastic waves beyond the diffraction limit. This phenomenon is called super-resolution and was seen previously in acoustic wave propagation.

4:55

1pSPb5. Imaging in random media. Liliana Borcea (Computational and Appl. Math, MS134, Rice Univ., 6100 Main St., Houston, TX 77005), James Berryman (Lawrence Livermore Natl. Labs., Livermore, CA 94551-9900), George Papanicolaou, and Chrysoula Tsogka (Dept. of Mathematics, Stanford Univ., Stanford, CA 94305)

We present a general method for estimating the location of small scatterers buried in a random, weakly scattering medium, using an active antenna array. The main features of this method are (i) an arrival time analysis of the echo received from the scatterers, (ii) a singular value decomposition of the array response matrix in the frequency domain, and (iii) the construction of an objective function in the time domain that is statistically stable and peaks on the scatterers. By statistically stable we mean that the objective function is self-averaging over different realizations of the medium. This is a new approach to array imaging that is motivated by recent studies of time reversal in a random medium. It combines features from seismic imaging like arrival time analysis with frequency domain radar signal subspace methodology like MUSIC (MUltiple SIgnal Classification).

MONDAY AFTERNOON, 3 DECEMBER 2001 ROOM 304, 1:00 TO 4:15 P.M.

Session 1pUW

Underwater Acoustics: Propagation Modeling and Environmental Effects II

Neil J. Williams, Chair

Division of Applied Marine Physics, RSMAS–University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149

Contributed Papers

1:00

1pUW1. Propagation model predictions and comparisons with measurements for broadband acoustic transmission in the Florida Straits. Harry DeFerrari, Neil Williams, and Hien Nguyen (Div. of Appl. Marine Phys., RSMAS–Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149)

Parabolic equation and normal mode propagation models are used to predict the channel pulse responses for comparison with experimental observations in the Straits of Florida. The model inputs for range dependent sound speed and bathymetry are known in detail but geoacoustic properties of the bottom are not well understood. In order to compensate, bottom input parameters are tuned so that predictions agreed with measurements. The tuned models are then used to explain the observed pulse arrival patterns and to predict other patterns that may come about from combinations of sound speed profiles and bottom properties that are known to exist and form interesting acoustic waveguides. Specifically, the models are used to examine the very intense focused arrivals from a deep source located in a downward refracting sound speed profile, the coupling of surface ducted and refracted modes to form ducted precursors and the possibility of strong attenuation of shipping noise with channel depth. Some of the predicted arrival patterns suggest experimental geometries that would enable ocean acoustic tomographic and geoacoustic inversions in shallow water.

1:15

1pUW2. Shallow water acoustic wave propagation experiments between three rigid bottom mounted acoustic sonar towers. Haruhiko Yamaoka and Tokuo Yamamoto (Div. of Appl. Marine Phys., RSMAS−Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149, hyamaoka@rsmas.miami.edu)

Shallow water acoustic wave propagation experiments were conducted between three bottom mounted 5-m tall towers as a part of the SWAT (Shallow Water Acoustic Technology) 2000 Experiment. Each sonar tower was self-contained to support a broadband sound transmitter and a hydrophone 5 m above seafloor. The Gold code of 1023 digits with a center frequency of 5.5 kHz was transmitted from each tower at 2- and 20-min intervals. The three towers were located in a triangle with ranges of 3, 4, and 5 km in the SS and OS sites. The results show a large difference in propagation characters. In the SS site, the peaks of arrivals are scattered over 10–15 ms from the first arrival peak. The envelopes of these signals are not stable during the 32-h experiment period. However, in the OS site the peak arrival is almost single within the 3-ms width. Also, the structure of arrival is very simple and includes the tidal effect. Parabolic equation (PE) model results based on CTD and bottom observation data indicate that there is a waveguide providing stable transmission just above the seafloor in the OS site. Comparison between acoustics data and broadband PE results will be discussed. [Work supported by ONR 321OA.]

1:30

1pUW3. Results from a shallow water propagation experiment on the Malta Plateau. Neil J. Williams, Harry A. DeFerrari, Hien B. Nguyen (Univ. of Miami, Rosenstiel School of Marine and Atmospheric Sci., 4600 Rickenbacker Cswy., Miami, FL 33149), Kurt Metzger (Univ. of Michigan, Ann Arbor, MI 48109-2122), Thomas H. Ensign, and Gary A. Zets (Engineering Acoustics, Inc., Winter Park, FL 32789)

A shallow water propagation experiment was performed in conjunction with a Saclant Undersea Research Center project known as the Malta Plateau Experiment (or MAPEX 2000) in early March 2000. A small, inexpensive, prototype 800-Hz source broadcasting 128-digit M-sequences was deployed off the southern coast of Sicily in approximately 135 m of water at a distance of 5 km from a fixed 64-element vertical array of hydrophones. This array relayed received signals to the R/V Alliance via radio telemetry. The Alliance then deployed a 256-element towed array and proceeded away from the source at a speed of 5 kts for about 3 h. The experiment was repeated once more using a second prototype with a different length M-sequence. The maximum distance from source to receiver was 30 km. The data were processed to remove Doppler and channel pulse responses were generated. The data will be compared with model results and the utility of the new projector will be discussed. [Work supported by ONR Code 321SS.]

1:45

1pUW4. Temporal coherence and time spread of ocean surface scattered high-frequency acoustic signals. Steven Lutz, David Bradley, R. L. Culver, Joe Keranen, and Jon Reeves (Appl. Res. Lab. and Grad. Prog. in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804)

The influence of the ocean environment on the temporal coherence and time spread of high-frequency, surface scattered acoustic signals over intervals ranging from a few seconds to tens of hours have been investigated experimentally. Pulses transmitted every 128 ms reveal persistent micropath arrivals subsequent to the main surface reflected signal. The observed structure has implications for channel equalization update rate requirements and optimal range resolution for target detection. The measurements were made during the (BBCoh/PasSAS) joint experiment with the Marine Physical Laboratory, Scripps Institution of Oceanography conducted in Fall 2000 off the coast of San Diego in water depths of 200 m to 700 m. Signals transmitted include 0.1-ms pure-tone pulses and 8-ms 13-kHz bandwidth linear frequency modulated (LFM) sweeps centered at 18, 32, and 46 kHz. The projector was suspended from the Floating Instrument Platform (FLIP) at a water depth of 170 m while the receive hydrophones were suspended from a moored buoy at two different ranges, 640 m and 2350 m, with water depths of 200 m and 700 m, respectively. Environmental measurements included water temperature, current speed and direction, directional wave height, and wind speed. [Work sponsored by ONR Code 321US under the ARL Program.]

2:00

1pUW5. Impact of depression to elevation internal soliton wave conversion on upslope acoustic propagation—A simulation. Marshall H. Orr and Peter C. Mignerey (Naval Res. Lab., 4555 Overlook Ave. S.W., Washington, DC 20375)

Depression to elevation soliton wave conversion has been observed at the South China Sea shelf break during the recent AsiaEX experiment. Numerical simulation of the internal soliton conversion process for a two-layer fluid has been used to generate a range/time-dependent sound speed profile for upslope propagation along a 20-km propagation path. Acoustic signal properties during various stages of soliton conversion are compared. [Work supported by the ONR Ocean Acoustic Program and ONRs NRL base funding.]

2:15

1pUW6. Validity conditions for the Markov approximation in theories for internal wave effects on propagation. Frank S. Henyey and Terry E. Ewart (Appled Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

There are two conditions that, if satisfied, guarantee the validity of the Markov approximation as usually imposed [see, for example, B. J. Uscinski *The Elements of Wave Propagation in Random Media* (McGraw-Hill, New York, 1977)]. One condition is that the scattering angle be small compared to the anisotropy of the medium, and the second is that there be less than one radian of phase shift in a correlation length. For the MATE experiment on acoustic propagation through oceanic internal waves, the second condition is violated, yet theories based on the Markov approximation work well. In this talk, it is shown that this second condition is much too pessimistic, by explicit construction of the next-order correction in an example that neglects sound channel effects over a correlation length. There is a strong cancellation among terms for the correction, each of which is of the order of the second expansion parameter squared. This cancellation occurs because the horizontal wave number is insensitive to the vertical wave number at small angles and that most of the phase shift comes from large-scale internal waves. With this cancellation, Markov approximation theories should correctly predict the MATE results—as they do. [Work supported by ONR.]

2:30

1pUW7. Effect of an extra-tropical hurricane on the underwater acoustic environment. Daniel Hutt, John Osler (Defence Res. Establishment Atlantic, 9 Grove St., Dartmouth, NS B2Y 3Z7, Canada), and James Abraham (Meteorological Services of Canada, Dorval, QC, Canada)

Hurricanes are a common occurrence along the east coast of North America in late summer and autumn. As a hurricane moves northward, its energy is sustained by circulation of the warm humid air over the Gulf Stream. Hurricanes usually lose energy quickly once they pass north of the Gulf Stream where they are affected by cooler surface water temperatures and mid-latitude winds. However, in some cases hurricanes may reintensify when the frontal zone encounters a strong upper-level trough. Hurricane Michael was one such extra-tropical hurricane that reintensified during passage along the Scotian Shelf in October 2000. Oceanographic conditions on the Scotian Shelf were monitored before, during, and after passage of the hurricane. Bathythermograph surveys and satellite imagery indicate that passage of the hurricane cooled surface waters and increased the thickness of the mixed layer. The effect of the hurricane on underwater acoustic propagation was investigated by calculating broadband transmission loss using a normal modes model in the 20-Hz to 10-kHz band. A comparison of acoustic transmission based on sound speed profiles from before and after Hurricane Michael and from climatology will be presented.

2:45

1pUW8. Shallow water impact acoustics. David M. Deveau (Naval Undersea Warfare Ctr. Det AUTEC, 801 Clematis St., West Palm Beach, FL 33401)

During a series of impact detection and tracking accuracy tests near Wallops Island, Maryland, an analysis was conducted on the acoustic impacts themselves. Since the impact area was very shallow (<60 ft.), the

acoustic signature traveled unscathed between the 1000 yard baseline sensors. The initial goal was to implement deep water impact tracking algorithms in the shallow environment. Initially the in-water tracking system computed positions that were close to those which were laser sighted—but not close enough. Since the transient acoustic signature was very evident, hence a stable timing mark, an investigation into the acoustics surrounding the impact was undertaken. A review of the data showed small transients which occurred distinctly before the large transient detection which had routinely been considered as the impact itself. Using a threshold detection scheme, a relative set of timing marks were generated based on these early transients and were found to produce a more accurate placement of the projectile impact. A further analysis consisted of reviews of controlled historical experiments that supported the experimental data and yielded a more precise impact detection scheme through a knowledge of the physical acoustics.

3:00–3:15 Break

3:15

1pUW9. Measurement and modeling of sound penetration into sandy sediment: Absolute pressure levels. Darrell R. Jackson, Kevin L. Williams, Eric I. Thorsos, and Steven G. Kargl (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

As part of the SAX99 high-frequency sediment acoustics experiment, a buried array was used to measure the pressure field in a sandy sediment with sources in the water column. The sediment was a medium sand with pronounced directional ripples. The buried array consisted of five vertical subarrays of six elements each spanning a horizontal aperture of approximately 30 cm and a vertical aperture of approximately 45 cm. Sources operating over the frequency range 11–50 kHz were moved to obtain various propagation directions relative to the ripple field with ranges and elevations that provided incident grazing angles both larger and smaller than the 30-deg critical angle. Signal levels measured by the buried hydrophones were compared with calculations based on a fluid sediment model and small-roughness perturbation theory, with both directional ripple and small-scale isotropic roughness. Under the assumption that hydrophone sensitivity in-sediment is the same as the value measured in water, data-model comparisons show similar azimuthal dependence and agree in absolute level to within about 2 dB for above-critical incidence and about 3 dB for below-critical incidence. [Work supported by ONR.]

3:30

1pUW10. An effective density model for acoustic propagation in sediments derived from the Biot theory. Kevin L. Williams (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, williams@apl.washington.edu)

An acoustic propagation model that approximates a porous medium as a fluid with a bulk modulus and effective density derived from the Biot theory will be presented. Within the framework of the Biot theory it is assumed that the porous medium has low values of frame bulk and shear moduli relative to the other moduli of the medium and these low values are approximated as zero. This leads to an effective density fluid model. It is shown that, for saturated sand sediments, the dispersion, transmission, reflection, and in-water backscattering predicted with this effective density fluid model are in close agreement with the predictions of the Biot theory. This agreement demonstrates that the frame bulk and shear moduli play only a minor role in determining several aspects of sand acoustics. Thus for many applications the effective density fluid model is an accurate alternative to full Biot theory and is much simpler to implement. [Work supported by ONR.]

3:45

1pUW11. An effective motionless medium for sound propagating through an inhomogeneous, moving fluid. Oleg A. Godin (CIRES, Univ. of Colorado at Boulder and NOAA/Environ. Technol. Lab. R/ET1, 325 Broadway, Boulder, CO 80305-3328, oleg.godin@noaa.gov)

Lord Rayleigh's idea of similarity between acoustic fields in a moving fluid and in a certain ''effective'' quiescent medium proved very helpful in understanding and modeling sound propagation in an ocean with currents and in an atmosphere with winds, as well as in other applications involving flows with small velocity compared to sound speed. Known as effective sound speed approximation, the idea is routinely utilized in the contexts of the ray theory, normal mode representation of the sound field, and the parabolic approximation. Despite the wide use of the concept of effective sound speed, no theoretical justification of Rayleigh's idea was published that would be independent of the chosen representation of the sound field and uniformly apply to distinct propagation regimes. In this paper, such a justification is provided by reducing boundary conditions and a wave equation governing sound fields in the inhomogeneous moving fluid with a slow flow to boundary conditions and a wave equation in a quiescent fluid with effective sound speed and density. The derivation provides insight into validity conditions of the concept of effective quiescent fluid. Conditions are discussed under which the effective parameters are insensitive to the wave propagation direction. [Work supported by ONR.]

4:00

1pUW12. A comparison of modal, beam, and PE modeled replicas for matched-field processing in the Santa Barbara Channel. Christopher O. Tiemann and Michael B. Porter (Sci. Applications Intl. Corp., 1299 Prospect St., Ste. 303, La Jolla, CA 92037, tiemannc@saic.com)

In order to evaluate the use of different propagation models in matched-field processing (MFP) applications, source localization was attempted using data from the Santa Barbara Channel Experiment and replicas generated by the adiabatic normal mode model KRAKEN, the beam model BELLHOP, and the parabolic equation model RAM. A conventional matched-field processor using all these models successfully located the source position in range and depth at both short (1 km) and intermediate (7 km) range, with marginal results at the longest 15 km range. Because signal-gain degradation varies with the model used, simulations were performed to quantify the degradation and relate that amount to a model mismatch, i.e., an error in array element localization.

NOTE: Attendance at this session requires payment of an additional registration fee.

MONDAY EVENING, 3 DECEMBER 2001 MARIOTT ATLANTIC BALLROOM III/IV, 7:00 TO 9:00 P.M.

Session 1eID

Interdisciplinary: Tutorial Lecture on Noise Propagation and Prediction Outdoors

Kenneth J. Plotkin, Chair
Wyle Laboratories, 2001 Jefferson Davis Highway, Suite 701, Arlington, Virginia 22202

Chair's Introduction—7:00

Invited Paper

7:05

1eID1. Noise propagation and prediction outdoors. Tony F. W. Embleton (80 Sheardown Dr., Box 786, Nobleton, ON L0G 1N0, Canada)

Concerns about noise in the community date back to the dawn of recorded history, have continued through Roman times, the Middle Ages, and are still with us. Many of the principal wave-propagation mechanisms involved in the real atmosphere outdoors were understood at least qualitatively by the late 1800s. Today, a good understanding of all the sound propagation phenomena is of considerable economic and social importance in connection with environmental impact studies. Reality is far more intesting than simple spherical spreading in the space above a hard, flat ground. Some grounds are hard like concrete or soft as snow. Corresponding reflection coefficients are usually less than unity and vary with angle. Grounds may not be flat, leading to shadow zones or alternatively multiple reflected paths. Wind convects sound waves, and gradients of wind or temperature refract waves either upward (temperature lapse) or downward (inversion). The atmosphere is rarely still and turbulence causes fluctuations in the acoustical effects. All these features are frequency dependent, and mutually interact. Measured sound levels outdoors owe as much to near-surface weather and to ground shape as to the obvious acoustical factors of source and receiver heights and their separation. The discussion will emphasize field measurements and simple physical interpretations.

Session 2aAB

Animal Bioacoustics: Fish Audition and Sound Production

Mardi C. Hastings, Chair
Biomedical Engineering Center, The Ohio State University, 1080 Carmack Road, Columbus, Ohio 43210

Chair's Introduction—8:00

Invited Papers

8:05

2aAB1. The enigma of fish ear diversity. Arthur N. Popper, Xiaohong Deng, John Ramcharitar, and Dennis M. Higgs (Dept. of Biol. and Neurosci. and Cognit. Sci. Prog., Univ. of Maryland, College Park, MD 20742, ap17@umail.umd.edu)

There are substantial interspecific differences in the gross anatomy of fish inner ears, including the relative size of the different end organs and the sizes and shapes of their otoliths and sensory epithelia. Differences also occur at finer structural levels and include the orientation patterns of the ciliary bundles on the sensory hair cells. There is also substantial structural variation along the lengths of a single sensory epithelium. This includes not only the mode of contact between the sensory epithelium and the otolith, but also the lengths of the ciliary bundles on the hair cells, and the ultrastructure of the hair cell bodies themselves. The functional significance of the interspecific differences in ear structure is far from being understood, but it may reflect the evolution of different ways to do the same basic kinds of peripheral signal processing, or different kinds of signal processing. We also do not understand the functional importance of the intraepithelial differences, although they may suggest variance in signal processing along the lengths of single epithelia. This paper will consider all of these differences, and suggest ways in which we might think about their functional significance.

8:30

2aAB2. Perception of sound by goldfish. Richard R. Fay (Parmly Hearing Inst., Loyola Univ. of Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626)

Psychophysical experiments have helped define the sense of hearing of goldfish with respect to detection and discrimination thresholds within many acoustic dimensions. Generally, these data reveal aspects of peripheral and early brain processing that are qualitatively, and sometimes quantitatively, indistinguishable from those of most other vertebrates. This correspondence suggests that goldfish share an essentially common sense of hearing with other vertebrates, and motivated new experiments that investigate more complex aspects of sound perception in goldfish using the phenomenon of stimulus generalization to study perceived similarities among simple and complex sounds. We have found that goldfish listen analytically to simultaneous tones, have perceptual dimensions analogous to pitch, timbre, and roughness in human listeners, and can segregate simultaneous, complex sources in perception. Several of these processes will be illustrated. These results are in accord with the hypothesis that the sense of hearing in goldfish is essentially similar to that revealed in other vertebrates studied. We suggest that it is useful to view the sense of hearing of any given vertebrate species as representing a general vertebrate pattern as an alternative to assuming that individual species are specially adapted to process particular sounds in its species-specific niche. [Work supported by NIH, NIDCD.]

8:55

2aAB3. Florida red tides and hearing. Zhongmin Lu and Zemin Xu (Dept. of Biol., Univ. of Miami, 1301 Memorial Dr., Rm. 4, Coral Gables, FL 33146, zlu@miami.edu)

Red tides, characterized by patches of colored seawater, have been reported worldwide. They are formed due to blooms of marine phytoplankton populations. Particularly, the dinoflagellate (*Ptychodiscus-brevis*) has been identified to cause the red tides occurring along the coastline of west Florida. It is known that *P. brevis* produces many types of neurotoxins, which result in massive kills of fishes, birds, and marine mammals including endangered species. Although some studies demonstrated effects of the neurotoxins on the respiratory system, heart, and neuromuscular junction, no study has ever been reported regarding possible effects of any red tide toxins on sensory systems. This study was to investigate whether or not a neurotoxin purified from Florida red tides, brevetoxin-3, affects hearing in fish. The neurotoxin was intraperitoneally injected into the goldfish (*Carassiusauratus*), and hearing thresholds were measured using auditory brainstem recording techniques. Results show that the toxin causes significant hearing loss in the goldfish. Species differences in neurotoxin tolerance and mechanisms of the hearing loss are discussed. Results from this study enhance our awareness of neurotoxin effects on sensory systems in vertebrates including humans. [Work supported by NIH Grant No. R29DC03275, UM start-up funds, and NIEHS pilot project award.]

9:20

2aAB4. Listening in the dark: Behavioral and neural mechanisms of acoustic recognition in a singing fish. Andrew H. Bass (Dept. of Neurobiology & Behavior, Cornell Univ., Ithaca, NY 14853)

Vocalizing species of teleost fish have a simple repertoire of species-typical signals central to their social and reproductive behavior. Male midshipman fish (''the California singing fish'') acoustically court females using ''hums'' that are long duration (secs 1h) and spectrally simple. Individual hums exhibit an essentially flat envelope consisting of a low fundamental frequency (90–100 Hz) and several harmonics. By contrast, naturally occurring acoustic beats show envelope modulations: the hums of neighboring males overlap to produce a beat with a difference frequency (0–8 Hz) in which amplitude and phase modulations result from the continual shifting of the two fundamental frequencies. Behavioral playback (phonotaxis) studies show that individual midshipman can use multiple signal parameters including hum duration, gap duration between successive hums, fundamental frequency, and intensity to distinguish hums from each other, while the effectiveness of beats as playback signals depends upon both beat modulation rates and depth of modulation. Neurophysiological evidence shows that auditory neurons in the midshipmans VIIIth nerve and midbrain encode the physical attributes of these signals mainly using spike rate and spike synchronization, mechanisms that we propose contribute to the neural templates adapted for both hum and beat recognition. [Research supported by NIH.]

9:45–10:05 Break

10:05

2aAB5. Using passive acoustics to monitor spawning of fishes in the drum family (Sciaenidae). Joseph Luczkovich (Inst. for Coastal and Marine Resources and Dept. of Biol., East Carolina Univ., Greenville, NC 27858) and Mark Sprague (Dept. of Phys., East Carolina Univ., Greenville, NC 27858)

The goal of this study was to develop a passive acoustic survey protocol to identify spawning habitats of sciaenid fishes. Based on comparisons of recordings of captive-fish and field recordings of species-specific courtship sounds, spawning areas of red drum, weakfish, spotted sea trout, and silver perch (Family Sciaenidae) have been identified in Pamlico Sound, NC. Sciaenid sounds were recorded using either a portable hydrophone deployed from a boat or timer-operated sonobuoys. Loudness of mating choruses of weakfish and silver perch was strongly correlated with the abundance of pelagic sciaenid-type eggs, suggesting that these areas were used for spawning. Sonobuoy recordings showed that weakfish spawned only in the high-salinity habitats near the inlets, spotted sea trout spawned predominantly in the low-salinity areas near the river mouths, silver perch spawned in both high- and low-salinity areas, and red drum spawned most commonly in low-salinity areas, but only in September. The loudness of mating choruses of silver perch was diminished when vocalizing bottlenose dolphins were present in the area around sonobuoys or when recordings of signature whistles were played back. Such ''acoustical avoidance'' behavior causes variation in sound pressure levels and could lead to lower estimates of abundance of spawning adults when conducting passive acoustic surveys.

10:30

2aAB6. Chorus behavior ecology in warm temperature sciaenid (*Pisces: Sciaenidae*) within a warm–temperate to tropical coastal lagoon. R. Grant Gilmore, Jr. (Dynamac Corp., Mail Code DYN-8, Kennedy Space Ctr., FL 32899, rggilmorej@aol.com)

Individual sound production and aggregative choral behaviors in sciaenid fishes of the Indian River Lagoon system of east central Florida have been under study since 1977. Long term studies of sonifery augmented with equally long term life history research has allowed classification of their call behaviors relative to their ecology. Five sciaenid species, silver perch, *Bairdiella chrysoura*, spotted seatrout, *Cynoscion nebulosus*, black drum, *Pogonias cromis*, red drum, *Sciaenops ocellata*, and weakfish, *Cynoscion regalis* can be divided into temporal, spatial, morphological, and vocal guilds. These guilds reveal partitioning of estuarine resources not only for habitat, food, and predation avoidance, but also for reproductive biology, sonifery, and larval-juvenile ecology. Variation in call type is significant between these species further presenting implications for co-evolved mating strategies under syntopic conditions within confined lagoonal ecosystems.

Contributed Papers

10:55

2aAB7. Pulsed sound production by Sciaenid fishes. Mark W. Sprague (Dept. of Phys., East Carolina Univ., Greenville, NC 27858, spraguem@mail.ecu.edu)

Many fishes, including members of the family Sciaenidae, produce pulses of sound by driving their swimbladders with sonic muscles. The swimbladder functions as a damped oscillator driven by the sonic muscles in a transient response. The radiated waveform contains information about the excitation function generated by the sonic muscles and the properties of the swimbladder. A theoretical model for pulsed sound production based on a driven damped oscillator [M. W. Sprague, J. Acoust. Soc. Am. **108**, 2430–2437 (2000)] can be used to establish relationships between sonic muscle excitation, swimbladder motion, and the radiated waveform. Analysis of sonic muscle excitations and corresponding waveforms produced by weakfish (*Cynoscion regalis*) (Sciaenidae) suggest that the swimbladder continues to oscillate or ''ring'' after the sonic muscle driving force stops. Portions of the waveform produced during this ringing of the swimbladder depend solely on swimbladder properties such as volume, which relates to fish size, and damping.

11:10

2aAB8. Psychoacoustics of ultrasound detection in clupeids. David A. Mann (College of Marine Sci., Univ. of South Florida, 140 7th Ave. S., St. Petersburg, FL 33701 and Mote Marine Lab., 1600 Ken Thompson Pkwy., Sarasota, FL 34236, dmann@marine.usf.edu), William N. Tavolga (Mote Marine Lab., 1600 Ken Thompson Pkwy., Sarasota, FL 34236), Dennis M. Higgs, and Arthur N. Popper (Dept. of Biol., Univ. of Maryland, College Park, MD 20742)

Recent behavioral and physiological studies have shown that some clupeid fishes are able to detect ultrasonic signals up to 180 kHz. Fishes in the subfamily Alosinae, including the American shad (*Alosa sapidissima*) and the gulf menhaden (*Brevoortia patronus*) are able to detect ultrasound. Yet other clupeids, including the scaled sardine (*Harengula jaguana*), Spanish sardine (*Sardinella aurita*), and the bay anchovy (*Anchoa mitch-*

illi), are not able to detect ultrasound. In this paper we review the ability of clupeids to detect ultrasound, and investigate the psychoacoustics of ultrasound detection, including frequency discrimination and masking by low-frequency and high-frequency sounds. [Work supported by NSF and NIH.]

11:25

2aAB9. Effect of swimbladder resonance on auditory bandwidth and sensitivity in teleost fishes. Mardi C. Hastings, Corrie J. Derenburger (Biomed. Eng. Ctr. and Dept. of Mech. Eng., Ohio State Univ., 1080 Carmack Rd., Columbus, OH 43210, hastings.6@osu.edu), and James J. Finneran (Sci. Applications Intl. Corp., Maritime Services Div., 3990 Old Town Ave., Ste. 105A, San Diego, CA 92110)

Experimental studies of four teleost species indicate that the response of the swimbladder to sound controls the auditory bandwidth of fish both with and without a direct connection between the swimbladder and the inner ear. Motion of the swimbladder in an acoustic field generated by an underwater loudspeaker was measured *in vivo* from 200 to 3000 Hz for channel catfish (*Ictalurus punctatus*), goldfish (*Carassius auratus*), American shad (*Alosa sapidissima*), and oscars (*Astronotus ocellatus*). Data were normalized with local acoustic energy density to obtain swimbladder frequency responses. Channel catfish and goldfish have two swimbladder chambers and Weberian ossicles that connect the anterior chamber to the inner ear, while American shad and oscars have a single swimbladder and no Weberian osscicles. (Shad do have auditory bullas; however, no motion of the bullas was detected.) Swimbladder frequency responses were compared with audiograms reported in the literature. Results indicate that auditory bandwidth correlates directly with the bandwidth of swimbladder resonance in all four species. Auditory threshold, however, does not correlate with amplitude of the swimbladder resonance. The two species with Weberian ossicles have lower auditory thresholds, but swimbladder resonances of approximately the same or somewhat lower amplitude than the other two species.

TUESDAY MORNING, 4 DECEMBER 2001 ROOM 305, 8:00 A.M. TO 12:00 NOON

Session 2aAO

Acoustical Oceanography and Physical Acoustics: Turbulence and Finestructure Studies I: Water Column Turbulence

Alex E. Hay, Chair
Department of Oceanography, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

Chair's Introduction—8:00

Invited Papers

8:05

2aAO1. Small-scale processes and vertical diffusion in the ocean. Ann E. Gargett (Ctr. for Coastal Physical Oceanogr., Old Dominion Univ., Norfolk, VA 23529, gargett@ccpo.odu.edu)

A depth profile taken anywhere in the ocean reveals large vertical gradients of oceanic water properties such as temperature, salinity, oxygen concentration, macronutrients such as nitrate and silicate, and micronutrients such as iron. These vertical gradients are formed at the very largest (depth/basin) scales of motion in the sea, and eradicated at the very smallest scales, those associated with three-dimensional turbulence. Turbulent enhancement of vertical diffusion over that associated with molecular motions is central to vital processes within the ocean. Vertical diffusion resupplies nutrients (the ''fertilizer'' for marine plant life) to the upper sunlit layers of the sea, underpinning marine food webs. The location and strength of turbulent diffusion determines the distribution of manmade pollutants introduced into the ocean, whether by design or by accident. More surprising perhaps is the essential role played by vertical diffusion in determining large-scale oceanic density structure, hence the crucial importance of small-scale turbulent processes to studies of natural and/or anthropogenic climate change. This presentation will review a variety of processes known or believed to generate the smallest scales of velocity and scalar variance in the sea, and describe the tools oceanographers presently use to measure and interprete salient characteristics of these small-scale processes.

9:05

2aAO2. The measurement of flow and turbulence levels using acoustical scintillation analysis. Daniela Di Iorio (Univ. of Georgia, 220 Marine Sci. Bldg., Athens, GA 30602)

For acoustical propagation in the coastal environment, small scale fluctuations in the temperature/salinity and flow velocity distorts the spherically spreading phase fronts which result in amplitude and phase fluctuations (scintillation). As the fine structure variability in the medium has a natural coherence time, it will be transported by the current. Detection of pattern motion is an essential concept exploited in acoustical scintillation and has been demonstrated in a variety of coastal channels. Important features of this approach are that the measurement is path averaged, is of the flow component perpendicular to the acoustic path and is sensitive to the Fresnel scale size which is chosen to lie within the inertial subrange of the turbulent random medium. Analysis of the signals goes beyond measurement of flow speed, and has included detection of turbulent velocity and scalar levels and turbulent anisotropy. For one-way propagation the turbulent kinetic energy dissipation rate is derived from the variance of the log-amplitude signal which is dependent on the level of the Kolmogorov spectrum dominated by current fluctuations. From reciprocal propagation, scalar and velocity turbulence levels are derived from accurate phase fluctuations. Detection of turbulent anisotropy is based on horizontal and vertical angle of arrival correlations using a square array with 1 m sides and thus sensitive to turbulent scale sizes approximately a few meters.

9:30

2aAO3. Acoustic measurements of global ocean shear. Robert Pinkel, Luc Rainville, Eric Slater, Lloyd Green, Mike Goldin, and Mai Bui (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0213)

The energetics of the ocean can be viewed as a cascade, with energy input at relatively large spatial scales, 1–10 000 km, and eventually dissipated by turbulent processes at centimeter–millimeter scales. The turbulence is thought to be triggered by the small-scale shear in the sea. Since the shear field serves as a mediator of ocean mixing, its dynamics are of central relevance to the overall mechanics of the ocean. To gather global shear data, a dual frequency Hydrographic Doppler Sonar System has been developed for the Scripps research vessel Roger Revelle. The System consists of a nested pair of Janus transducers operating at 50 and 140 kHz. The 50 kHz sonar profiles to depths of 700–1000 m with 15 m depth resolution. The 140 kHz sonar achieves ranges of 200–300 m with 3 m depth resolution. Operating since January 2000, the system has detected considerable geographic variability in both shear and acoustic scattering strength. The deep Sea of Japan and the Eastern North Pacific are surprisingly low in shear. Large shears are encountered near the Hawaiian Ridge and in the waters beneath the Kuroshio. With continued operation, the large-scale patterns and underlying governing principles of the shear field will hopefully emerge.

9:55

2aAO4. Acoustic measurement of oceanic turbulent kinetic energy dissipation rate, using a large-eddy technique. Ann E. Gargett (Ctr. for Coastal Physical Oceanogr., Old Dominion Univ., Norfolk, VA 23529, gargett@ccpo.odu.edu)

Turbulence in the ocean results from many different processes operating over a wide range of time and space scales, with spatial and temporal variability particularly extreme in coastal oceans. If origins and effects of turbulent processes in the ocean are to be understood, it is essential to supplement the small number of accurate measurements of turbulent properties produced by current microprofiler techniques. This talk will discuss a new acoustic large-eddy technique for estimating turbulent kinetic energy dissipation rate ϵ, using measurements of vertical velocity from a specialized Acoustic Doppler Current Profiler (ADCP). The method was calibrated against a subset of ϵ profiles taken during a single cruise with now-standard airfoil probes carried on a free-fall microprofiler. Without further adjustment of the calibration constant, the large-eddy acoustic estimates successfully describe the remaining profiler data from the same cruise, as well as data obtained with a different microscale profiler on a subsequent cruise. The new technique reveals the complexity of the field of ϵ in energetic coastal tidal fronts, on spatial and temporal scales which cannot be resolved by standard microprofiler measurements.

10:20–10:35 Break

Contributed Papers

10:35

2aAO5. Observation of small scale fluid processes in the vicinity of continental shelf breaks using high-frequency acoustic backscattering systems. Marshall H. Orr (The Naval Res. Lab., 4555 Overlook Ave. S.W., Washington, DC 20375)

Acoustic signal vertical and horizontal coherence properties are controlled on the shelf and near the shelf break by a variety of flow induced fluid processes. These processes perturb the sound speed field at spatial scales of a few hundred meters and less. The frequency of occurrence, depth distribution, and the relationship of the sound speed field variability to both oceanic and atmospheric fluid processes occurring on the continental shelf is poorly understood and are not generally included in most acoustic propagation models. Acoustic flow visualization records collected on the continental shelf and at the shelf break in water depths ranging from 50–300 m will be presented. These records will show: (1) the types of internal wave and smaller scale turbulence that occur in the summer and early fall in the vicinity of the New Jersey shelf; (2) some observation of the dissipation of large amplitude internal waves at the shelf break of the South China Sea; and (3) a variety of small scale instabilities that can perturb the sound speed field. [Work supported by the ONR Ocean Acoustic Program and ONRs NRL base funding.]

10:50

2aAO6. Laboratory measurements of acoustic scattering from a temperature and salinity gradient. Joseph D. Warren,[a] Andone C. Lavery, Ray W. Schmitt, and Timothy K. Stanton (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543, jwarren@whoi.edu)

Recently developed theoretical scattering models predict that acoustic scattering from temperature and salinity microstructure at high frequencies (10 kHz–1 MHz) can be significant under certain oceanographic conditions. The results of this theoretical work suggest that it may be possible to use acoustic scattering techniques, in combination with the scattering models, to estimate oceanographic parameters, such as the dissipation rate of turbulent kinetic energy. In addition, quantification of the scattering contributions from temperature and salinity microstructure can play an important role in correctly interpreting acoustic surveys of marine life, since recent field data indicate that the acoustic returns from zooplankton and microstructure can be of similar strength. Acoustic backscatter from a sharp temperature and salinity gradient was measured in a tank capable of generating, and sustaining, a thin double-diffusive layer (1–2 cm thick) between two water masses (cold, fresh residing above warm, salty water). Vertical shear, temperature, and salinity profiles were measured during the experiment to provide input to the acoustic scattering models. Backscatter was measured at frequencies between 24 kHz and 500 kHz, and as a function of range from the sharp interface as part of a program to measure the backscattering from microstructure. [a] Currently at Southwest Fisheries Science Center, La Jolla, CA 92037.

11:05

2aAO7. Acoustic scatter from thermal plumes. John Oeschger (Coastal Systems Station, 6703 W. Hwy. 98, Panama City, FL 32407)

Laboratory results of high-frequency acoustic scattering measurements from a thermally driven plume are reported. Farfield weak scattering theory describes the acoustic scatter from the medium variability. To first order, the dominant scattering mechanism is thermally driven sound speed variations that are related to temperature deviations from ambient. As a result, the received complex acoustic scatter is a measure of the one-component three-dimensional Fourier transform of the temperature difference field measured at the Bragg wave number; this is the Fourier component of the plume variability that produces scattering. Results are presented for multistatic scattering from unstable and turbulent plumes using a parallel scattering geometry. Through common Bragg wave number comparisons, the data justify the application of the farfield weak scattering theory to the present case of a thermal plume. Results of temperature time series measurements taken of the plumes are also shown. Acoustical and mechanical temperature spectral estimates are presented and show a near-classical $-\frac{11}{3}$ power law behavior.

11:20

2aAO8. Measured turbulence in ship wakes. R. Lee Culver and David Bradley (Grad. Prog. in Acoust. and Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804)

Large surface ships moving at high speed generate extremely turbulent velocity fields containing high concentrations of air bubbles that are distributed in a nonuniform manner. The bubbles are acoustically active, making sonar a useful tool for imaging the ship wake and measuring the spatial and temporal variations of the bubbles. We have employed a 250-kHz multibeam sonar, integrated into an autonomous vehicle (AUV), to image the wakes of two large surface ships from the underside. The AUVs traveled at high speed so that the spatial variation of wake of approximately the same age could be assessed, and ran under different age sections of the wake in order to observe temporal variation. Using an assumed bubble size spectrum for the wake, bubble density has been estimated from the backscattered signals. Variations in bubble density with depth, cross-wake, and along-wake dimension, and with time, are observed. Spatial wave number spectra are used to quantify the scales over which bubbles are distributed, which may be related to the scales of turbulent energy distribution. Comparing measurements of different age wakes indicate how the distribution over spatial scales evolves with time. The multibeam sonar appears to provide quantitative information about the distribution of bubbles, and possibly turbulence, in the ship wake. [Work supported by ONR.]

11:35–12:00

Panel Discussion

TUESDAY MORNING, 4 DECEMBER 2001 ROOM 223, 8:00 A.M. TO 12:30 P.M.

Session 2aBB

Biomedical Ultrasound/Bioresponse to Vibration: Topical Meeting on Physics of Ultrasound in Relation to the Biology of its Therapeutic Effects: HIFU, LOFU, and Imaging HIFU Lesions

Pierre D. Mourad, Chair
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

Chair's Introduction—8:00

Invited Papers

8:05

2aBB1. MRI guided focused ultrasound opening of blood brain barrier. K. Hynynen, N. McDannold, N. Vykhodtseva, N. Sheikov, and F. Jolesz (Dept. of Radiol., Brigham and Women's Hospital and Harvard Med. School, 75 Francis St., Boston, MA 02115)

Previously, we demonstrated that focused ultrasound can focally open the blood brain barrier (BBB) reversibly without damage to the neurons. In this study, we varied the ultrasound pulse parameters to investigate their influence on BBB opening and any resulting tissue damage. Under magnetic resonance imaging (MRI) guidance, a series of targeted ultrasound bursts (1.5 MHz, burst duration: 5 cycles to 100 ms, repetition frequency: 1 Hz–1 kHz) were delivered into rabbit brains at low pressure levels (up to approximately 5 MPa) after intravenous administration of a commercially available ultrasound contrast agent containing preformed gas bubbles. Focal MRI signal intensity changes were observed during the deposition of acoustic energy. BBB opening at the correct target was confirmed by detecting the localized uptake of an MRI contrast agent. In some animals Trypan Blue was injected intravenously after the sonications. The animals were sacrificed at time intervals ranging from 2 h to 7 days. Light and electron microscopy evaluation was performed. No neuronal damage was observed at the lowest power levels that produced BBB opening. No neuronal damage was observed at the lowest power levels that produced BBB opening. Correlation of focal MRI signal changes to the uptake of a large molecular size contrast agent was observed demonstrating the potential of image guided drug delivery.

8:25

2aBB2. Dual-mode HIFU phased arrays for noninvasive surgery. Emad Ebbini (Dept. of Elec. and Computational Eng., Univ. of Minnesota, Minneapolis, MN 55455)

Ultrasound phased arrays employing advanced piezocomposite transducer technology are now commonplace in high-power therapeutic applications. When properly designed, these arrays can produce (single or multiple) therapeutic foci suitable for noninvasive surgery for a variety of applications. A properly designed HIFU array has a characteristic therapeutic operating field (ThxOF) around its geometric center. ThxOF defines the volume within which the array is capable of producing therapeutic effects at the intended focal spot(s) without collateral damage at other locations. Recently, we were able to show that HIFU arrays optimized for therapy has a characteristic imaging field of view (IxFOV) around its geometric center. IxFOV defines the volume within which the array is capable of producing images of the target at a specified contrast level. In this talk we outline the optimization procedure for designing dual-mode HIFU arrays suitable for imaging and therapy. In addition, we give illustrative experimental examples of dual-mode operation of HIFU phased arrays. In particular, we emphasize new imaging techniques for visualization of HIFU lesions, both discrete and volumetric. Images of HIFU lesions obtained by the dual-mode HIFU array and a commercial scanner will be compared to demonstrate the superiority of our specialized imaging modes in lesion visualization.

8:45

2aBB3. Bubbles and acoustic image-guided high intensity focused ultrasound. Michael R. Bailey, Shahram Vaezy, Jonathan C. Yuen, Ajay Anand, Nathan A. Miller, Peter J. Kaczkowski, and Lawrence A. Crum (Ctr. for Industrial and Medical Ultrasound, APL/UW, 1013 NE 40th St., Seattle, WA 98105)

Clinical diagnostic ultrasound (US) can be used to target and to monitor in real-time high-intensity focused ultrasound (HIFU) therapy. In our system, the HIFU transducer (3.5 MHz, 35 mm aperture, 55 mm radius of curvature) and US scan head (several were tested, center frequencies 3–8 MHz) are fixed with the HIFU focus in the imaging plane. HIFU and US are either synchronized real time to relegate interference to the image fringe or HIFU and US are interlaced for nearly real-time imaging. HIFU produces a localized hyperechoic region visible on B-mode US. Coagulatively necrosed lesions produced have similar size, shape, and location to measurements made from the corresponding US images. Thresholds are also comparable. However, *in vivo*, if HIFU is turned off as soon as hyperecho appears, no lesion is seen (the tissue was fixed within four hours of treatment). Thus, a short HIFU burst can be used to target treatment. Bubbles appear to be largely but perhaps not entirely responsible for the increase in echogenicity. Times for dissipation of the hyperecho and dissolution of a bubble as a function of hydrostatic pressure compare well. Significant overpressure (50 bar) can suppress hyperecho produced by HIFU. [Work supported by NSF, NSBRI, DARPA/ONR, and NIH SBIR.]

9:05

2aBB4. High intensity focused ultrasound for arrest of bleeding. Shahram Vaezy, Roy Martin, and Lawrence Crum (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98195)

High Intensity Focused Ultrasound (HIFU) has been shown to provide an effective method of hemostasis, in animal studies, for both solid organs and blood vessels (acoustic hemostasis). Moderate to profuse bleeding from major blood vessels, liver, and spleen, can be stopped within 1–2 min of HIFU application. The efficacy of acoustic hemostasis has been demonstrated using HIFU frequencies of 1–10 MHz, and acoustic intensities of 1000–5000 W/cm^2. Conventional B-mode ultrasound imaging and Doppler appear to provide a effective visualization, targeting, and monitoring methods when extracorporeal application, in case of internal bleeding, is desired. Both thermal and mechanical mechanisms of HIFU appear to be responsible for achieving hemostasis. The thermal mechanism, due to ultrasound absorption, raises the tissue temperature in excess of 70 °C in less than a second, leading to shrinkage of the tissue and collapse of small blood vessels. The mechanical mechanism, due to cavitation activity, disrupts the tissue structure, leading to the occlusion of large blood vessels. Both mechanisms result in coagulative necrosis of tissue, and arrest of bleeding. Acoustic hemostasis may provide an effective method in surgery and prehospital settings for treating bleeding in trauma and elective surgery patients.

9:25

2aBB5. On the increase of disintegration efficiency of an acoustic lithotripter. V. R. Singh (Dept. of Biomed. Measurements and Standards, Natl. Physical Lab., New Delhi-110012, India, vrs@csnpl.ren.nic.in)

Currently, disintegration of renal calculi, noninvasively, with the use of lithotripters, is in common use. In the hospitals, commercial lithotripters like extra-corporeal shock wave lithotripters (ESWL) are in routine use. These systems are very cumbersome and expensive. Treatment of the patients also takes comparatively more time, due to more number of sittings. The high ultrasonic intensity used in such systems has side effects, and causes damage to delicate nerves and fibers in the surroundings of the stones present in the kidney. Therefore, the present lithotripters are required to be improved by increasing the efficiency of disintegration and other factors. In the present work, the increase in the disintegration efficiency of the ultrasonic lithotripters, in this case Siemens ESWL, is made by stimulating externally the main shock wave with an acoustic stress wave. The cavitation bubbles are found to implode faster, with more disintegration efficiency of the lithotripter, giving in turn, better treatment to the patient, with less sitting time and less ultrasound dosage level.

9:45–9:55 Break

9:55

2aBB6. High intensity focused ultrasound treatment of abdominal tumors. G. ter Haar, I. Rivens, and M. Allen (Dept. of Phys., Inst. of Cancer Res., Royal Marsden Hospital, Sutton, Surrey SM2 5PT, UK)

A phase II clinical trial for the treatment of isolated liver metastases is currently underway at the Royal Marsden Hospital. Tumors lying 4–12 cm below the skin surface are targeted in fully conscious patients. Treatment success is monitored using either radiological methods in conjunction with biochemical marker assays, or by histological examination. Factors determining pain or other sensations have been studied in order that exposures may be optimized. Laboratory investigations have been undertaken in parallel with these clinical studies, and these have been used to establish the effects of, for example, rib shadowing, timing of individual exposures, and *in situ* intensity on tissue ablation. Skin studies have allowed the determination of thresholds for a ''burn'' when multiple, overlapping exposures are used.

10:15

2aBB7. Transrectal focused ultrasound treatment of localized prostate cancer. Jean-Yves Chapelon, RTmi Souchon, Laura Curiel, and Albert Gelet (Inserm U556, 151 cours A. Thomas, 69424 Lyon, France)

Since 1989, a research project with the aim of establishing treatment of localized prostate cancer by means of high intensity focused ultrasound (HIFU) has been developed. The treatment of localized prostate cancer with HIFU, is administered transrectally, using ultrasound guidance only. The quality of the induced necrosis depends to a large extent on the choice of firing parameters: power, frequency, numbers and position of shots, shots duration, waiting time between shots and the shape of the acoustical field. The

lesions are created by a thermal effect. Their slightly conical form is due to the absorption of the ultrasound wave by the tissues, increased by cavitation bubbles. Recently, it has been shown that real time elastography has the potential to accurately depict the ablated volume during a HIFU session, therefore providing feedback on the treatment. Results obtained since 1993 demonstrate that transrectal HIFU prostate ablation is an effective therapeutic alternative for patients with localized prostatic adenocarcinoma.

10:35

2aBB8. Ultrasonic monitoring of radiation-force deformation induced by therapeutic ultrasound. Frederic L. Lizzi, Robert Muratore, Cheri X. Deng, Samuel Mikaelian, S. Kaisar Alam, and Jeffrey A. Ketterling (Riverside Res. Inst., 156 William St., New York, NY 10038-2609)

We have developed a dual-transducer technique to monitor tissue deformation induced by therapeutic ultrasound. The technique is designed to monitor and control therapeutic lesion size and location. Currently, the technique employs a spherical-cap high-intensity transducer with a central, co-linear diagnostic transducer. A-mode rf echo signals are first acquired using the diagnostic transducer to document pre-exposure conditions within the target tissue. A short (10 ms) intense exposure is then delivered with the spherical-cap transducer in order to induce tissue motion by radiation force. A-mode rf signals are then acquired for 0.2 s. Pre- and post-exposure A-mode signals are analyzed using correlation processing to determine the extent of induced tissue motion. The time constant associated with return to initial tissue positions is also calculated. These procedures have been applied *in vitro* to liver specimens using the spherical-cap transducer to periodically induce motion and to produce thermal lesions; at relatively low exposure levels, results are consistent with computer simulations. Larger exposures can produce permanent A-mode signal changes. We are optimizing this technique to enable it to validate therapy beam placement, to evaluate induced compression of vessels, and to monitor tissue rigidity changes that accompany thermal-lesion formation.

10:55

2aBB9. SonablateTM 500: A novel ultrasound platform for transrectal image-guided treatment of prostate diseases. Narendra T. Sanghvi, Roy F. Calson, Ralf Seip, Jahangir Tavakkoli (1 Focus Surgery, Inc., Indianapolis, IN 46226), Kris A. Dines (XDATA Corp., Indianapolis, IN 46220), Thomas A. Gardner (Indiana Univ. Med. School, Indianapolis, IN 46202), and Toyoaki Uchida (Kitasato Univ. Med. School, Sagamihara, Japan)

The SonablateTM system for transrectal image-guided high intensity focused ultrasound (TR-HIFU) treatment of prostate diseases has been in clinical use since 1992. Recently, we developed the third generation of the system, call SonablateTM 500, with major improvements aimed specifically toward localized prostate cancer treatment. Salient features of the SonablateTM 500 include: (1) PC-based digital technology, (2) multifocal lengths probe with split beam transducer to cover a wide range of tissue depths (2.5–4.5 cm), (3) an upgraded treatment planning and control software with safety features such as reflectivity index measurement (RIMTM) and transducer tissue measurement (TTMTM); and (4) 3-D volume image rendering of the prostate. The SonablateTM 500 platform was tested in an *in vivo* canine experimental study in which the entire prostate was ablated. The prostate ultrasound images before and post HIFU treatment show very good correlation to the planned lesion. Histology delineated that the entire prostate was contiguously ablated without affecting surrounding tissues. Long-term canine survival studies as well as human prostate cancer studies under the FDA's IDE (Investigational Device Exemption) approval are in progress. [We would like to thank the engineering and clinical staff of Focus Surgery, Inc. who made this project possible.]

11:15

2aBB10. HIFU for tumor therapy in China: A report on its application to humans for a variety of diseases. Feng Wu and Wang Zhibiao (Box 153, Inst. of Ultrasound Eng. in Medicine, Chongqing Medical Univ., 1 Medical College Rd., Chongqing 400016, PROC)

Starting in 1997, our center began to apply HIFU in clinical treatment to patients with malignant solid tumor, such as breast cancer, hepatic carcinoma, and bone sarcoma, etc. During three years of study, we have observed satisfactory results from more than 400 cases, including some patients with bone sarcomas. A brief description of our treatment methodology will be reviewed as well as a few individual cases.

11:35–12:30
Panel Discussion

Session 2aEA

Engineering Acoustics: Small Scale Acoustics and Acoustic Devices

Daniel M. Warren, Chair
Knowles Electronics, 1151 Maplewood Drive, Itasca, Illinois 60143

Chair's Introduction—8:15

Invited Papers

8:20

2aEA1. A silicon nitride microphone diaphragm inspired by the ears of the parasitoid fly *Ormia ochracea*. Ronald N. Miles, Colum Gibbons, Jia Gao (Dept. of Mech. Eng., State Univ. of New York, Binghamton, NY 13902-6000, miles@binghamton.edu), Kyutae Yoo, Quang Su (Cornell Univ., Ithaca, NY 14853), and Weili Cui (State Univ. of New York, Binghamton, NY 13902-6000)

A differential microphone is described that has been designed to employ similar operating principles to that of the ears of the parasitoid fly, *Ormia ochracea*. The ears of this fly have been shown to be highly directional even though they are only about 1 mm across [R. N. Miles, D. Robert, and R. R. Hoy, J. Acoust. Soc. Am. **98**, 3059–3070 (1995)]. Analyses of the mechanics of this biological system suggest novel approaches to the design of miniature directional microphones. Finite element analysis results for the acoustic resonse of a 1 mm by 2 mm silicon nitride microphone diaphragm are presented. The diaphragm responds to pressure gradients in a manner that is inspired by *Ormia's* ears. Predicted results for the natural frequencies, mode shapes, frequency response and directivity of our design are shown to compare closely with measured data obtained for a prototype silicon nitride diaphragm. [Work supported by NIH and DARPA.]

8:40

2aEA2. Wireless telephone design considerations. Robert A. Zurek (Personal Communs. Sector Res. and Adv. Technol. Labs, Motorola, 2001 N. Division St., Harvard, IL 60033, w15754@email.mot.com)

As wireless communication devices have shifted from novelty devices to everyday use, the demands on the performance and quality of the device's audio and acoustic systems has increased. The acoustic portion of wireless phone design can no longer be taken for granted as the size of the devices decrease and the expectations increase. While wireless telephones share many of the design challenges of wired landline systems, they also present additional challenges not normally found in landline systems. The shape and size of today's phones make past assumptions used in the design of landline systems invalid, such as the assumption that the acoustic load on the receiver is constant. The delay inherent in today's digital systems coupled with the small size of handsets has led to an increased awareness of acoustic feedback from the receiver to the microphone, which results in an undesirable system echo. In addition to the receiver design and echo concerns is the necessity of designing in an alert or ringer that must produce adequate SPL levels to inform the user of a call, while fitting in a very small space. The aforementioned issues will be discussed along with current design techniques utilized to overcome these challenges.

9:00

2aEA3. Scaling issues in a submillimeter MEMS microphone. Peter V. Loeppert (Emkay Innovative Products, 1151 Maplewood Dr., Itasca, IL 60143)

A silicon based microphone with a 0.5 mm diam diaphragm has been developed using a combination of surface and bulk micromachining or MicroElectro Mechanical Systems (MEMS) techniques. This device is intended for high volume general purpose markets such as telephony, where cost is a key factor. In batch fabricated MEMS devices, cost is critically determined by the device size. Yet the performance standards for this microphone are set by relatively large electret condenser microphones (ECM) built in competing technologies. This paper will discuss the scaling issues involved in meeting high performance goals in a very small MEMS device. In particular, tradeoffs made to achieve high sensitivity, low noise, and unconditional stability will be discussed.

9:20

2aEA4. Turbulence induced distortion in hearing aid receivers. Daniel M. Warren (Knowles Electronics, LLC, 1151 Maplewood Dr., Itasca, IL 60143, daniel.warren@knowles.com)

The quality of resonances in hearing aid receivers can be reduced by adding damping in various places in the electroacoustic system. Acoustic damping may be implemented by means of tiny ($\approx$0.001 in. diam) holes in the diaphragm separating the front and rear cavities of the receiver. The holes are very small, $\approx$1/10 of the boundary layer thickness. As noted previously [P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, 1968)], ''fortunately, such a minute hole is of no practical

interest.'' However, in hearing aid receivers such holes are of practical interest, and are identified as causing high distortion at frequencies below 500 Hz. The experiment described demonstrates that the distortion is caused by turbulence in air flow through holes in the diaphragm. The experiment also shows that distributing the air flow among multiple smaller holes raises the threshold for turbulent flow, alleviating distortion at the nominal drive level. [Work supported by Knowles Electronics, LLC.]

Contributed Papers

9:40

2aEA5. Challenges for the commercialization of acoustic sensors using microfabrication technologies. Michael Pedersen (CNRI, 1895 Preston White Dr., Ste. 100, Reston, VA 20191, pedersen@mems-exchange.org)

In this paper a review is given of the methods used in the design and realization of acoustic sensors based on microfabrication technologies. It is shown that the most important issues relate to material properties, which must be tightly controlled to achieve acceptable consistency and stability of the devices. In addition, the requirements for successful commercialization of such devices are discussed and a number of critical parameters are identified which include environmental performance, reliability and cost. It is contended that microphones made with this new technology will not provide advances over existing technology in terms of sensitivity and noise performance, and hence the motivation for the use of the technology must be found in the added value of smaller size, increased bandwidth, and better flexibility in terms of device assembly and ruggedness. Based on these considerations a new capacitive microphone design is proposed, which simplifies the fabrication of the so-called free-plate or zero-stress microphone, referring to the lack of tension in the microphone diaphragm. It is further shown that the introduction of flip-chip assembly methods will enhance the repeatability and ruggedness of the microphone package, which so far has received little attention in the development micromachined acoustic sensors.

9:55–10:10 Break

10:10

2aEA6. FEM based simulation of nonlinear effects in armature alerts. Robert A. Zurek and Simone Koo (Personal Communs. Sector Res. and Adv. Technol. Labs, Motorola, 2001 N. Division St., Harvard, IL 60033, w15754@email.mot.com)

Although alerts or ringers have been used in wireless phones and pagers since their inception, design of these systems in the past has been limited to estimation of the Helmholtz resonance of the acoustic system followed by trial and error modification until the desired alert level is achieved. With the advent of reasonably fast Vibro-Acoustic Boundary Element and Finite Element solvers, a much more accurate and inclusive method of alert system design is possible. BEM and FEM alert models can take into account the baffling effects of the phone or pager case, losses due to surface absorption and transmission loss, geometrical characteristics of the system's ports and cavities, and the frequency dependent vibration characteristics of the transducer. These simulation capabilities, combined with Doppler laser vibrometer based alert characterization, provide for an improved alert design cycle. This paper will discuss some of the new alert characterization and simulation techniques used in the Motorola PCS Research Advanced Products Lab. The results of this technique will be compared to measured data and earlier simulation attempts, and future enhancements to the current technique will also be discussed.

10:25

2aEA7. New materials for thermal microphone sensors. Jerry Helffrich and Heather Hanson (Southwest Res. Inst., 6220 Culebra Rd., Bldg. 178, San Antonio, TX 78238-5166, jhelffrich@swri.org)

Existing thermal microphone sensors use MEMS technology to fabricate two beams. These initially heated beams cool with the movement of air in a sound wave. The temperature difference between the two beams due to convective heat transfer can be converted to an air velocity measurement. These microphones have the advantage of an extremely small size and dipole sensing, but suffer from a low signal-to-noise ratio, low-frequency response, and high-power dissipation. We are investigating the use of new, high-temperature coefficient of resistivity (TCR) materials for thermal microphones. This second generation of devices should deliver an improved response than those that are commercially available. Measurements of the response of thermal microphones using new resistive materials will be presented.

10:40

2aEA8. Thermoelastically actuated MEMS ultrasonic transducer. Venkataraman Chandrasekaran (Dept. of Aerosp. Eng., Mech., and Eng. Sci., Univ. of Florida, Box 116250, Gainesville, FL 32611), Xi Li, Jian Li, Toshikazu Nishida (Dept. of Elec. and Computer Eng., Univ. of Florida, Box 116130, Gainesville, FL 32611), Louis N. CattafestaIII, and Mark Sheplak (Dept. of Aerosp. Eng., Mech., and Eng. Sci., Univ. of Florida, Box 116250, Gainesville, FL 32611)

Ultrasound based techniques provide a method for noncontact distance measurement utilizing a thin vibrating diaphragm as the acoustic source. Devices based on different operating principles have been reported including electrostatic and piezoelectric. The conventional pulse-echo technique is limited by a finite pulse decay time that restricts the measurement range to greater than a few centimeters. Proximity sensing on the order of millimeters can be achieved using continuous wave form techniques such as the acoustic impedance method or simultaneous excitation at multiple frequencies. Using silicon microfabrication technology, a MEMS ultrasonic transducer has been fabricated for millimeter range proximity sensing. The salient features of the sensor/actuator include a 0.01 mm thick, 1 mm diameter composite diaphragm that is thermoelastically driven at ultrasonic frequencies using a circular heater at its center. Dynamic Joulean heating of the diffused surface heater induces bending moments in the diaphragm due to the temperature gradient across the membrane cross section. Electromechanical transduction of the membrane vibration is achieved via four diffused silicon piezoresistive strain gauges located at the edge of the diaphragm. The design, fabrication and preliminary characterization of this sensor/actuator are presented.

10:55

2aEA9. Measurement of particulate densities in air. Ravinder Beniwal and Sean Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

Clean air is one of the most important issues that govern the health of all live forms. However, presently there are not many quick and simple methods for measuring impurities like particulates in air. These impurities have an enormous diversity in their physical and chemical structure. They may be unburned carbon particles from a diesel engine exhaust and chimney, pollen grains in the spring air, or asbestos in a factory. This paper shows that changes in the composition of the air cause a change in the speed of sound. Therefore by measuring the change in the speed of sound, it is possible to monitor the density of particulates in the air. Preliminary tests are conducted on various smoke–air mixtures. The results demonstrate that this methodology is very sensitive to any changes in the composition of the air. Its implementation is very simple and efficient, and costs much less than the conventional method currently used in the auto industry. This technique will be used to calculate the mass density of the particulates resulting from a diesel engine, and results thus obtained will be compared with those calculated using other methods.

11:10

2aEA10. Capacitive micromachined interdigital transducers. Goksen G. Yaralioglu, Mohammed H. Badi, and Butrus T. Khuri-Yakub (Ginzton Lab., Stanford Univ., Stanford, CA 94305)

Capacitive Micromachined Ultrasonic Transducers (CMUTs) have become very popular for medical imaging and sonar applications during the last decade. They can compete with piezoelectric transducers in terms of efficiency and bandwidth. Recently, it has been noticed that these devices couple energy into silicon wafer in addition to the immersion medium. By placing the CMUTs on a substrate in an interdigitated configuration, Lamb wave or Rayleigh wave modes can be excited with very high efficiency without a need for any piezoelectric material. In this study, the acoustic power coupled into the silicon substrate as well as the acoustic impedance of the CMUT membrane are calculated by using finite element analysis. Normal mode theory is used to find the distribution of the acoustic power among different Lamb wave modes. For low frequency (1 MHz) devices, the lowest order antisymmetric (A_0) mode Lamb wave is the dominant mode in the substrate. For high frequency devices (100 MHz), interdigital CMUTs excite Rayleigh wave with efficiencies comparable to piezoelectric surface acoustic wave (SAW) devices. Experimental results for Lamb wave devices in the vicinity of 1 MHz will also be presented. [Work supported by ONR.]

11:25

2aEA11. Difference method for measuring the vibration sensitivity of microphones. Anthony J. Brammer (Inst. for Microstructural Sci., Natl. Res. Council, Montreal Rd., Ottawa, ON K1A 0R6, Canada)

In some applications, microphones are subjected to sufficient vibration that their vibration response is of interest (e.g., feedback from earphone to microphone in wireless telephones, and hearing aids). A procedure for measuring this microphone property should take into account the airborne sound produced by the apparatus. In the present method, two microphones of the same type are mounted facing each other with diaphragms parallel, and form the plane, circular ends of a cylindrical enclosure. The dimensions of the air-filled volume between the two microphones are chosen to ensure that both microphones experience the same sound pressure. When the enclosure is vibrated in a direction normal to the plane of the microphone diaphragms (i.e., along the long axis of the cylinder), the vibration responses of the two microphones are essentially out of phase. The difference signal between the two microphone outputs will then provide a measure of the mean vibration response of the microphones from which the common airborne response of the microphones has been subtracted. The extent to which the response to sound within the apparatus interferes with the measurement of the vibration response of individual microphones may then be estimated. Results will be presented for a miniature electrical microphone.

TUESDAY MORNING, 4 DECEMBER 2001 BALLROOM B, 10:00 A.M. TO 12:00 NOON

Session 2aED

Education in Acoustics: Hands-on Demonstrations for High School Students

Uwe J. Hansen, Chair
Department of Physics, Indiana State University, Terre Haute, Indiana 47809

Chair's Introduction—10:00

This session is intended to communicate the excitement of ''doing'' acoustics to high school students in the South Florida area. Individual demonstrations with associated instrumentation will be presented in a brief introduction. In the subsequent hands-on session, presenters will guide individual students to discover the fun of acoustics by letting them ''twiddle the knobs.'' Topics available for student participation will range in sophistication from simple wave observation to normal mode mapping on musical instruments. While attendance as well as participation by conference attenders is welcome, hands-on participation is primarily designed for the students attending this session by special arrangement.

Session 2aMU

Musical Acoustics: Musical Instrument Acoustics

James P. Cottingham, Chair
Physics Department, Coe College, Cedar Rapids, Iowa 52402

Contributed Papers

8:00

2aMU1. Shape optimization techniques for musical instrument design. Luis Henrique (Instituto Politécnico do Porto, Escola Superior de Música e das Artes do Espectáculo, R. da Alegria, 503, 4000-046 Porto, Portugal), José Antunes (Instituto Tecnológico e Nuclear, Appl. Dynam. Lab., 2686 Sacavém Codex, Portugal), and João Soeiro de Carvalho (Universidade Nova de Lisboa, Faculdade de Ciências Sociais e Humanas, Av. de Berna 26-C, 1050-069 Lisboa, Portugal)

The design of musical instruments is still mostly based on empirical knowledge and costly experimentation. One interesting improvement is the shape optimization of resonating components, given a number of constraints (allowed parameter ranges, shape smoothness, etc.), so that vibrations occur at specified modal frequencies. Each admissible geometrical configuration generates an error between computed eigenfrequencies and the target set. Typically, error surfaces present many local minima, corresponding to suboptimal designs. This difficulty can be overcome using global optimization techniques, such as simulated annealing. However these methods are greedy, concerning the number of function evaluations required. Thus, computational times can be unacceptable if complex problems, such as bell optimization, are tackled. Those issues are addressed in this paper, and a method for improving optimization procedures is proposed. Instead of using the (local) parameters $P_f(\mathbf{r})$ as searched variables, the system geometry is modeled in terms of truncated series of orthogonal space-functions $P_j(\mathbf{r})=\Sigma_n a_{jn}\psi_{jn}(\mathbf{r})$, and optimization is performed on the (global) coefficients a_{jn}. Fourier series and orthogonal polynomials are typical functions $\psi_{jn}(\mathbf{r})$. This technique reduces considerably the number of search variables, and has a potential for significant computational savings in complex problems. It is illustrated by optimizing the shapes of both current and uncommon marimba bars.

8:15

2aMU2. Violin normal mode radiative properties. George Bissinger and John Keiffer (Phys. Dept., East Carolina Univ., Greenville, NC 27858)

Calibrated far-field pressure measurements of violin radiation excited by force hammer impact at the G-string corner of the bridge were performed at 266 microphone positions over an $r=1.2$ m spherical surface inside an anechoic chamber. These were combined with simultaneous mobility measurements over the violin surface using a scanning laser to extract radiation efficiencies for violin normal modes below 4000 Hz. For favorable cases it was possible to compute experimental values for radiation damping, and the companion fraction of vibrational energy lost to acoustic radiation. As an example, preliminary results for the lower of the first corpus bending modes of a violin by A. Schroetter were the following: radiation efficiency of ~ 0.12, total damping of ~ 1.15 (percent critical) for a Q of ~ 43, radiation damping of ~ 0.41 (percent critical). For this mode $\sim 35\%$ of the vibrational energy was radiated as sound. Radiation patterns measured for each mode were corrected for anechoic chamber reflection/interference effects. [Research supported by the NSF DMR-9802656.]

8:30

2aMU3. How important are torsion modes in bowed-string dynamics? Octávio Inácio (Instituto Politécnico do Porto, Escola Superior de Música e das Artes do Espectáculo, R. da Alegria, 4000-046 Porto, Portugal), José Antunes (Instituto Tecnológico e Nuclear, Appl. Dynam. Lab., 2686 Sacavém Codex, Portugal), and Luis Henrique (Instituto Politécnico do Porto, Escola Superior de Música e das Artes do Espectáculo, R. da Alegria, 503, 4000-046 Porto, Portugal)

In spite of the venerable history of bowed-string research, several aspects are still inadequately clarified. The importance of torsion modes on the motion regimes of bowed strings is one such issue. Experiments involving torsion are difficult to perform and most of the results available pertain to numerical simulations. The authors' approach differs from previous efforts in two main aspects: (1) the development of a computational method quite distinct from the classic wave-propagation approach pioneered by McIntyre, Schumacher, and Woodhouse and (2) an extensive and systematic analysis of the coupling between torsion and transverse motions. The numerical simulations are based on a modal representation and a friction model that enable accurate representations of the stick-slip friction forces and of the string dynamics, in both time and space. Many relevant aspects of the bowed-string can be readily implemented, including string inharmonic behavior, finite bow-width and torsion effects. Concerning the later aspect, a realistic range of the wave-speed ratio $\alpha \equiv C_{Tors}/C_{Tran}=2\sim 7$ is investigated, for several values of the bow velocity and normal force. The results suggest that torsion modes can affect both transient durations and steady state regimes, in particular when $\alpha < 4$. Gut strings should then be particularly prone to torsion effects.

8:45

2aMU4. Comparisons of theoretical and experimental transverse motion for a plucked, stiff piano string. José Sánchez and James Irwin, Jr. (Dept. of Elec. and Computer Eng., Bradley Univ., 1501 W. Bradley Ave., Peoria, IL 61625)

This work is meant to determine whether central or backward difference approximations in computer simulations more closely approximate reality. The stiff piano string motion is experimentally measured by a real-time analog optical technique that was developed to measure the one-dimensional transverse displacement of a plucked string. The optical technique uses a detector that collects laser light not obstructed by the string. The received light is converted to a voltage, which is related to the string displacement perpendicular to the laser beam. The signal is then filtered and conditioned by external circuitry, digitized by a *Mona* data acquisition system, and are displayed using *Cool Edit Pro*. These measurements are then compared to two different finite difference computer simulations. One of the computer simulations uses central differences and the other uses backward differences, both with the appropriate boundary conditions and parameters. The comparison of the experimental data and the computer simulations is comprised of informal listening tests, spectral, and time domain comparisons.

9:00

2aMU5. The HANG: A hand played steel drum. Thomas D. Rossing (Phys. Dept., Northern Illinois Univ., DeKalb, IL 60115), Uwe J. Hansen (Indiana State Univ., IN), Felix Rohner, and Sabina Schärer (PanArt, Bern, Switzerland)

The HANG is a new steel instrument consisting of two spherical shells of steel, suitable for playing with the hands. Seven to nine notes are harmonically tuned around a central deep note which is formed by the Helmholtz resonance. Elliptical note areas are formed on the upper shell and tuned so that the $(1,1)_a$ and $(1,1)_b$ modes of each note area have frequencies that are in the ratios of 3:2:1 with respect to the fundamental, except in the highest notes where they are in the ratio of 2.5:2:1 or 2.4:2:1. Holographic studies show the modal shapes and also the extent of coupling between note areas.

9:15

2aMU6. Acoustic reflectometry in the manufacture of reproductions of historical wind instruments. D. Murray Campbell, James M. Buick, Jonathan A. Kemp, and Maarten O. van Walstijn (Dept. of Phys. and Astron., Univ. of Edinburgh, James Clerk Maxwell Bldg., Edinburgh EH9 3JZ, U.K., D.M.Campbell@ed.ac.uk)

Acoustic reflectometry has become a well-established technique for investigating the internal bore profile of cylindrically symmetrical ducts. Although the method has previously been applied with some success to the study of musical wind instrument bores, there are practical problems which limit the accuracy of bore reconstruction of instruments which are long or widely flaring, or which have side finger holes. This paper reports on a collaboration with Christopher Monk Instruments Ltd., a London-based firm specializing in the reproduction of cornetts and serpents. These are lip-reed instruments with side finger holes, and a curved bore center line which inhibits the use of conventional bore measuring techniques. It is demonstrated that acoustic reflectometry can provide bore profiles of these instruments with sufficient precision to be a valuable guide in their construction or repair. The precautions necessary to achieve this level of accuracy are described, and prospects for further refinements are discussed. [The collaboration of Keith Rogers and Jeremy West of CMI Ltd. is gratefully acknowledged. Work supported by the Engineering and Physical Sciences Research Council, U.K.]

9:30

2aMU7. Acoustical analysis of an ancient Mexican flute. Rolando Menchaca (CIC, IPN, Mexico), Sergio Beristáin (Lab. Acustica, ESIME, IPN, Mexico and and P.O. Box 75805, 07300, Mexico City, Mexico), and Roberto Velazquez (CIC, IPN, Mexico)

The acoustical anslysis made on one flute taken from a given class among very old Mexican flutes is discussed. First, some acoustical characteristics of the sounds produced from this musical instrument are presented, such as the acoustic power and the energy spectrum of some of the different sounds generated by the flute. Then an analysis is made on the acoustical details of the construction technology used to build this kind of instrument, like acoustical mass and compliance, turbulence generation, radiation from the tuning holes, and so on. And some discussion on the probable design objectives on the mind of the ancient Mexicans when they developed and built this sound instrumentation, as the possibility that natural sounds were taken as a reference for their art evolution.

9:45

2aMU8. On the acoustics of organ reed pipes. Eric Cox and Thomas D. Rossing (Dept. of Phys., Northern Illinois Univ., DeKalb, IL 60115)

In most lingual organ pipes, the reed vibrates against a fixed shallot and modulates the flow of air passing through the shallot into the resonator. In an earlier paper [T. D. Rossing, J. Angster, and A. Miklos, paper 2pMU6, 136th meeting, Acoustical Society, 1998] we described distinct regimes of oscillation observed in a Trompette pipe. Using a laser vibrometer, we have now observed similar phenomena in two pipes of quite different design: a Rankett and a Krummhorn. Without the resonator the reed frequency in each pipe varies smoothly with reed length, but the resonator lowers the frequency of the reed and also produces distinct regimes of sound, separated by discontinuities, as the system jumps from one regime to the next one. Changing the wind pressure has a modest effect on sound output but little effect on sounding frequency, within limits.

TUESDAY MORNING, 4 DECEMBER 2001 ROOMS 118/119, 8:30 TO 11:50 A.M.

Session 2aNS

Noise, Architectural Acoustics and Signal Processing in Acoustics: Implementing the Classroom Acoustics Standard

John Erdreich, Chair

Ostergaard Acoustical Associates, 200 Executive Drive, West Orange, New Jersey 07052

Chair's Introduction—8:30

Invited Papers

8:35

2aNS1. Acoustical environments measured in urban and suburban schools. John Erdreich (Ostergaard Acoust. Assoc., 200 Executive Dr., West Orange, NJ 07052)

Measurements of sound levels in numerous classrooms and other instructional spaces in primary educational facilities over the past several years have revealed a lack of acoustical control within public schools. We have found environments in excess of NC-60 more often than we have found NC-30 rooms in unoccupied spaces. Occupied sound levels in instructional spaces may exceed 90

dB(A). Because of poor acoustical design, we have found classrooms on the same floor with background sound levels that vary from NC-60 to NC-25, depending on the distance from the HVAC mechanical room. This paper will present some case studies that demonstrate the need for an acoustical standard for the classroom.

9:05

2aNS2. Development of an American National Standard for classroom acoustics. David Lubman (D. Lubman & Assoc., 14301 Middletown Ln., Westminster, CA 92683-4514) and Louis C. Sutherland (Acoust. Consultant, 27803 Longhill Dr., Rancho Palos Verdes, CA 90725-3908)

A draft ANSI (American National Standards Institute) Standard for classroom acoustics will be submitted for approval this year. It is believed that adopting this Standard will result in a more inclusive and effective education system. Although eliminating acoustical barriers benefits everyone, young children and persons with hearing, language, speech, and learning disabilities are special beneficiaries. Titled ''Acoustical performance criteria, design requirements, and guidelines for schools,'' the draft, designated ANSI S12.60-200X, limits noise and reverberation in classrooms and imposes minimum values of sound isolation between enclosed classrooms and adjacent spaces. The effort began when many members of the Acoustical Society of America (ASA) became convinced that noise and reverberation in classrooms were serious barriers to learning. The ASA organized a Task Force on classroom acoustics, assembling volunteer scientists and engineers from architectural acoustics, noise, speech, and psychological acoustics. Task Force outreach seminars led to a broad coalition of individuals, and public and private organizations. They determined that a national Standard for classroom acoustics was needed. To develop that Standard, ANSI Working Group 42 on classroom acoustics was established and scoped. Draft 10 was successfully balloted in the S12 Committee in January 2001. In this paper we discuss the development of the Standard.

9:35

2aNS3. Experiences with acoustical renovation of classrooms in a public school. Robert Wallace (The Harbor Science and Arts Charter School, New York, NY 10029; robertw104@aol.com)

The Harbor Science and Arts Charter School is a small, public elementary school with children in grades 1 through 7 located in the East Harlem section of New York City. A hallmark of our curriculum is that we encourage students to work on group projects and to make many oral presentations of the results of their work. We noticed that when children were engaged in group work and oral presentations, it was often difficult to understand everything that was said. One of our fourth grade teachers had a friend who was an acoustical designer who arranged a sound analysis of the teacher's room that confirmed that our high ceilings, concrete block walls, and window air conditioners were creating some of the problems we were experiencing. Using the room as a test case, Armstrong Industries donated materials for modifications to the room. They subsidized lowering the ceiling with acoustical tiles. We installed new bulletin boards and purchased new air conditioners. The results were extraordinary. Students and teachers who used the classroom were amazed at the improvements. We have used what we have learned to modify all of the classrooms as well as the cafeteria. These relatively simple modifications have made a significant improvement in the quality of life in our school.

10:05–10:20 Break

10:20

2aNS4. Implementation of ANSI S12.60 acoustical performance criteria, design requirements and guidelines for schools. Lois E. L. Thibault (Coordinator of Res., U.S. Access Board, 1331 F St. NW, #1000, Washington, DC 20004-1111, thibault@access-board.gov)

In 1997, the parent of a child with a hearing loss petitioned the U.S. Access Board to include guidelines for classroom acoustics in the accessibility guidelines the Board maintains under the ADA. The Board agreed that poor listening conditions in schools could be a barrier to the education of children with hearing impairments and other disabilities and arranged to collaborate with ASA and other stakeholders on an acoustical standard for classrooms. The Board will submit the completed standard to the International Codes Council for reference in the International Building Code adopted by many states and local jurisdictions, making it enforceable through the local permitting and inspections process. Reference in the ADA Accessibility Guidelines may follow in its next review cycle in 2005. Individualized Education Plans (IEPs) required under IDEA (Individuals with Disabilities Education Act) may also reference the new standard.

10:50

2aNS5. The new standard's HVAC background noise requirements and their implementation in building design. Mark E. Schaffer (Schaffer Acoust., Inc., 869 Via de La Paz, Ste. A, Pacific Palisades, CA 90272) and Jerry G. Lilly (JGL Acoust., Inc.)

The new ANSI classroom standard recommends long-term, steady-state background sound level limits of 35 dBA and 55 dBC. Since the HVAC system is typically a room's dominant steady-state noise source, it is usually reasonable to apply these limits to the design of the system, assuming that other noise sources produce insignificant additions to the room's long-term background noise environment. The recommended sound level limits are significantly lower than current industry standards, and complying with them will increase system costs. School administrators, architects, and HVAC system designers should, therefore, be prepared for the changes. This paper summarizes equipment selection and system design guidelines that have been found to be necessary for ensuring compliance with the sound level limits.

11:20

2aNS6. The Future of Schools, Inside and Out: TV documentary. Debra Greene and Crystal Edgington (Information Television Network, 621 NW 53rd St., Boca Raton, FL 33487)

''The Future of Schools, Inside and Out'' is an information-rich and colorful 1 hour documentary designed to entertain and inform viewers about America's schools. Areas of interest include acoustics, building safety, energy efficiency, day lighting, and more. The episode explores the connection between our nation's school facilities and the quality of a child's education. This program travels out in the field to school locations from rural areas to the inner city to the suburbs to help tell the story. Several top experts not only point out the problems and challenges but also offer solutions to help give the children of America every possible advantage in their learning environment.

TUESDAY MORNING, 4 DECEMBER 2001 ROOM 222, 8:00 TO 11:15 A.M.

Session 2aPA

Physical Acoustics: High Amplitude Effects in Resonators I

Anthony A. Atchley, Cochair
Graduate Program in Acoustics, Pennsylvania State University, P.O. Box 30, Applied Science Building, State College, Pennsylvania 16804

Victor W. Sparrow, Cochair
Graduate Program in Acoustics, Pennsylvania State University, P.O. Box 30, Applied Science Building, State College, Pennyslvania 16804

Invited Papers

8:00

2aPA1. Flow and temperature fields in thermoacoustic stacks. Ph. Blanc-Benon and S. Duffourd (Ecole Centrale de Lyon, Ctr. Acoustique, LMFA UMR CNRS 5509, BP 163 69131 Ecully Cedex, France, Philippe.Blanc-Benon@ec-lyon.fr)

The paper describes the flow patterns and the temperature distribution measured in a thermoacoustic engine. Experiments have been performed in a resonant standing wave thermoacoustic refrigerator model. The knowledge of temperature and flow fields in the microchannels and at the edges of the stack plates becomes an increasingly important issue in the design of heat exchangers for thermoacoustic engines. Different geometries of the stack plates have been investigated as well as at different pressure levels. Heat exchangers have also been placed on both sides of the stack, with adjustable gaps. 2-D particle image velocimetry was used to investigate the flow with enlarged observation fields near the ends of the stack plates. We present velocity profiles across the microchannels, 2-D velocity fields. Vorticity maps were deduced using a criterion based on a normalized angular momentum. Vortex motion close to the stack ends has been clearly observed, however, with a strong influence of the plate thickness. An array of miniaturized thermocouples permits us to follow in real time the building-up of the temperature gradient along the stack for different oscillating flow conditions and different gaps between the stack and the heat exchangers. Finally, comparisons are made with theoretical models and numerical simulations recently published in the literature. [Work supported by DGA.]

8:30

2aPA2. Numerical simulation of viscous losses in resonators. Philip J. Morris and Said Boluriaan (Dept. of Aerosp. Eng., Pennsylvania State Univ., 233 Hammond Bldg., University Park, PA 16802)

Thermoacoustic devices consist of resonators that contain high amplitude standing or traveling waves. They depend on the interaction between the high amplitude acoustic waves with surfaces in the resonator, such as a stack, to either transport heat or extract power. However, this same interaction results in losses. These losses may be in the form of viscous dissipation or from the conversion of acoustic energy into vorticity. The latter occurs at sudden changes in the cross sectional area of the resonator where the unsteady flow produces a jet- and sinklike behavior in one period of the oscillation. Numerical simulations are performed on a parallel cluster of PC's for the unsteady flow generated at such a sudden change of area. Results are presented for the instantaneous and averaged flow field and the thermodynamic properties. The pressure drop across the sudden change in area is calculated and the physical processes that result in this pressure difference are examined. Estimates are given of the total losses in the resonator and the relative contributions of viscous dissipation in the wall boundary layers and the conversion of acoustic energy into vorticity are given. [Research supported by ONR.]

9:00

2aPA3. Minor losses in oscillatory flow. Barton L. Smith and Gregory W. Swift (Condensed Matter and Thermal Phys. Group, Los Alamos Natl. Labs., Los Alamos, NM 87545)

Energy dissipation that is manifested by gradients in pressure in a flow system is often expressed in terms of a pressure loss factor of the form $K=\Delta 2P/\rho v^2$. While there is a wealth of data on loss factors for a myriad of geometries in steady flow, little such data exists for pulsatile or oscillatory flow. Recent advances in thermoacoustics, which involves heat engines and refrigerators with no moving parts, motivate work in this area. An oscillatory flow system has been constructed to investigate loss factors in 2D geometries. The current results provide loss factors of entrances/sudden expansions. It is found that the losses are primarily a function of the displacement amplitude. The results also provide information on the variation of the ''equivalent length'' with the same parameter, and on turbulent dissipation in the channel. Oscillating flow in diffusers is also investigated, with emphasis on loss factors as a function of diffuser angle, and unsteady separation effects. [Work supported by the Los Alamos National Laboratory LDRD funds.]

9:30–9:45 Break

Contributed Papers

9:45

2aPA4. Measurements of acoustic Einstein–Hopf drag on objects undergoing volume oscillations. Edward Tucholski and Andres Larraza (Naval Postgrad. School, Phys. Dept., Monterey, CA 93943, larraza@physics.nps.mil)

It has been shown theoretically that the drag on an object undergoing volume oscillations can be modified by the presence of isotropic, homogeneous, broadband acoustic noise, when the band overlaps the objects resonance width [Larraza and Tucholski, Phys. Rev. Lett. **84**, 2378–2380 (2000)]. These results constitute an acoustic analog to the Einstein–Hopf drag on an oscillating dipole in the presence of electromagnetic fluctuations. However, an important difference is that band-limited acoustic noise can reduce the drag when the lower frequency of the spectrum coincides with the resonant frequency of the object. Experimental results of the drag measured on bubbles and aerogel spheres moving in an band-limited noise field are reported.

10:00

2aPA5. Liquid droplet generation utilizing high intensity acoustic standing waves within shaped resonators. Josh Roby, Tim Cameron, and Bart Lipkens (Dept. of Mech. Eng., Virginia Commonwealth Univ., 601 W. Main St., Richmond, VA 23284-3015 and Dept. of Psych., Boston Univ., Boston, MA 02115)

Generating liquid droplets from a bulk liquid underlies a vast number of process related technologies ranging from fuel combustion to respiratory drug delivery. Resonant Macrosonic Synthesis (RMS) is a technology that makes possible the generation of very high-amplitude shock-free acoustic standing waves within shaped resonators. Sound pressure levels in excess of 190 dB have been measured within RMS resonators. Bulk liquids subject to these high intensity oscillating acoustic waves are drawn into the acoustic field and atomized. Under continuous processing conditions water droplets on the order of 30 μm Sauter Mean Diameter have been generated with RMS. Droplet size measurements were obtained using a laser diffraction particle-sizing instrument. Data indicate that under all but the lowest liquid flow rates, liquid accumulates within the acoustic resonator, altering the processing characteristics. As liquid accumulation increases under fixed drive amplitude the power delivery, acoustic pressure, and acoustic resonance frequency drop, which coincides with increased mean particle diameters. Therefore the mean droplet diameters increase with an increasing liquid flow rate. Experiments have shown, however, that these characteristics are highly dependent on the process entry and removal points on the resonator.

10:15

2aPA6. Simulation of streaming in a tapered resonator. Brian C. Tuttle and Victor W. Sparrow (Grad. Prog. in Acoust., Penn State, 217 Appl. Sci. Bldg., University Park, PA 16802)

In 1997, Olson and Swift [Cryogenics **37**, 769–776 (1997)] investigated acoustic streaming in conical tubes of varying taper angle of a pulse tube refrigerator to increase the efficiency. They found that an optimal taper angle existed that would minimize Rayleigh streaming in the tube. The present research is aimed at numerically simulating this taper effect on the Rayleigh streaming. The two-dimensional (2-D) nonlinear acoustic numerical approach of Sparrow and Raspet [J. Acoust. Soc. Am. **90**, 2683–2691 (1991)] is modified for tapered geometries matching those of Olson and Swift. The method for grid generation and stretching will be discussed. Results of this numerical simulation will be compared to the published experimental results and will aid in a deeper understanding of acoustic streaming. [Work supported by ONR.]

10:30

2aPA7. Application of a stream-function numerical approach to Rayleigh streaming. Debbie Sastrapradja and Victor W. Sparrow (Grad. Prog. in Acoust., Penn State, 217 Appl. Sci. Bldg., University Park, PA 16802)

In 1995 Kamakura *et al.* [J. Acoust. Soc. Am. **97** (5, Pt. 1), 2740–2746] showed how an incompressible, viscous fluid stream-function vorticity formulation could be applied to model the time dependent build up of acoustic streaming in and around the focal area of an ultrasonic transducer in water. They found excellent qualitative agreement between their numerical results and experimental findings available in the literature. The purpose of the present research is to apply this numerical method to the Rayleigh streaming encountered in resonator tubes of thermoacoustic devices. A new code has been written from scratch and benchmarked against the published results of Kamakura *et al.* with excellent agreement. An overview of both the algorithm and application method to the Rayleigh streaming problem briefly will be provided. Initial results and comparisons to recent analytical solutions will be given. While not as flexible as high-resolution computational fluid dynamics calculations on large parallel computers for the study of acoustic streaming, the present formulation is much faster and can run on a desktop machine. [Work supported by ONR.]

10:45

2aPA8. Acoustic streaming generated by standing waves in narrow channels. Mark F. Hamilton, Yurii A. Ilinskii, and Evgenia A. Zabolotskaya (Dept. of Mech. Eng., Univ. of Texas, Austin, TX 78712-1063)

Considerable attention has been devoted to acoustic streaming following the initial investigation by Rayleigh. A standing wave in a channel generates acoustic streaming in the form of inner and outer vortices that circulate in opposite directions. The outer vortex is the streaming pattern described by Rayleigh. The inner vortex is generated within the viscous boundary layer and has comparable thickness. Because boundary layer thickness is typically an exceedingly small fraction of channel width, the inner vortex is normally disregarded in relation to the Rayleigh streaming. However, channel widths in stacks used in thermoacoustic engines are comparable to boundary layer thickness. As channel width decreases, the relative thickness, of the inner vortex increases as that of the outer vortex decreases, and when channel width is comparable to the boundary layer thickness, the inner vortex dominates the flow field. Previous studies of inner streaming vortices have focused mainly on the problem of a cylinder oscillating in an infinite fluid. Acoustic streaming in narrow channels is studied here analytically and numerically. An analytical solution is derived in the quadratic approximation. Fully nonlinear numerical results are obtained from a 2-D code recently developed for thermoacoustic engines. Analytical and numerical results are in good agreement. [Work supported by ONR.]

11:00

2aPA9. Numerical study of resonant gas oscillation in a closed tube with varying cross section. Takeru Yano (Dept. of Mech. Sci., Hokkaido Univ., Sapporo 060-8628, Japan, yano@mech-me.eng.hokudai.ac.jp)

Nonlinear resonant gas oscillations in closed tubes with varying cross-sections filled with an ideal gas are numerically studied. The resonant gas oscillation concerned is excited by a piston at one end of the tube, which oscillates harmonically with a resonant frequency obtained by solving the quasi-1-D linear wave equation. The system of Navier–Stokes equations for axisymmetric flow is directly solved by a high-resolution upwind TVD finite-difference scheme. The numerical results may be summarized as follows: (i) If the variation of cross-section is small, a quasi-steady oscillation state includes shock waves, as in the tubes of uniform cross-section. (ii) When the variation of cross-section is relatively large, the gas oscillation evolves into a standing wave of large amplitude, but shock waves are not formed. (iii) The (nondimensional) maximum amplitude can attain the order of the cubic root of M, where M is the acoustic Mach number at the sound source. (iv) Such a large amplitude oscillation induces relatively fast acoustic streaming.

TUESDAY MORNING, 4 DECEMBER 2001 ROOM 221, 9:00 TO 11:30 A.M.

Session 2aSA

Structural Acoustics and Vibration: Methods of Passive Dissipation in Structural Acoustics I

George Lesieutre, Chair

Department of Aerospace Engineering, Pennsylvania State University, State College, Pennsylvania 16802

Invited Papers

9:00

2aSA1. Passive damping applications for noise and vibration reduction. Conor D. Johnson, Roger M. Glaese, and Paul C. Janzen (CSA Eng., Inc., 2565 Leghorn St., Mountain View, CA 94043)

This paper presents an overview of several unique techniques and their applications for passive damping for vibration and noise control. Most of the methods that are used target structural vibration and by its attenuation, transmitted acoustics as well. Viscoelastic materials are the most widely used technology for applications. An innovative family of viscoelastic materials, having fairly constant modulus and loss factor properties with respect to temperature, has recently been developed, with some materials in this family having very good thermal conductive properties as well. Another innovative viscoelastic method involves co-curing the viscoelastic and a composite host structure. Additionally, an innovative acoustic enclosure for precision manufacturing has been developed that relies on barrier mass and structural damping to attenuate acoustic transmission. Passive magnetic damping has been used in both tuned mass dampers and passive isolation systems for structural-acoustic systems. Research has been conducted on the design and applications of particle damping. By placing a number of small particles in a cavity of a vibrating structure, passive damping may be achieved over wide frequency and temperature ranges. The use of all of these methods will be illustrated through ground, air, and space-based examples.

9:30

2aSA2. A method for increasing the damping properties of fiber reinforced composite structures. Anthony A. Caiazzo and B. Walter Rosen (Mater. Sci. Corp., 500 Office Ctr. Dr., Ste. 250, Fort Washington, PA 19034)

The passive material damping concept discussed in this paper is a sandwich construction of continuous fiber reinforced composite layers separated by a constrained visco-elastic material (VEM) layer. The fiber orientations in the composite layers are spatially oriented to produce systematic stiffness variations. It has been shown that three independent shear components may be developed in the VEM for simple applied loads using this spatially varying stiffness (SVS) approach, which is a distinct improvement over other traditional constrained layer treatments. To demonstrate predicted high-damping levels, modal loss factor data are presented for simple tubular and beam bending elements. Results of airborne sound transmission loss measurements on 1-m square SVS panels of various areal densities are described. Random vibration levels for an equipment mounting plate constructed using the SVS material are also presented and compared to results for a metallic base plate.

10:00

2aSA3. Analytical/experimental studies of integral passive damping in composite isogrid structures. Ronald Gibson and Yu Chen (Mech. Eng. Dept., Wayne State Univ., Detroit, MI 48202)

Composite isogrid structures are currently in use in a number of aerospace applications where the control of structural vibration response is an important design consideration. In this presentation, a recent investigation of the modal vibration characteristics of composite isogrid structures with embedded viscoelastic damping layers will be summarized. Analysis was done by means of a finite element method implementation of the modal strain energy method, while experiments were done by using an impulse-frequency response technique. Composite isogrid test panels were fabricated by a fast and inexpensive thermoplastic stamping process using two different co-mingled materials: glass/polypropylene and carbon/nylon. The damping materials were embedded at the interface between the ribs and the skin where the shear deformations are greatest in order to significantly improve the damping properties of the structures. It is found that the required amount of damping material can be reduced significantly if it is optimally placed. The modal vibration characteristics of composite isogrid structures are also compared with those of other composite structures such as laminates and sandwich panels. The authors gratefully acknowledge the support of the WSU Institute for Manufacturing Research, as well as donations of materials by Vetrotex America, Cytec Fiberite, and 3M Company.

10:30–10:45 Break

Contributed Papers

10:45

2aSA4. Damping measurements of highly damped wavy composite structures. Scott D. Sommerfeldt (Dept. of Phys., Brigham Young Univ., Provo, UT 84602), William F. Pratt, and Matthew S. Allen (Patterned Fiber Composites, Inc., Lindon, UT 84042)

Wavy composite material is a composite that has been developed in recent years that uses opposing sinusoidal fiber waveforms in adjacent layers in a manner that enhances the constrained layer damping effect in the material. Analysis of the material has predicted both high damping and stiffness, which is contrary to many materials where increased damping generally leads to decreased stiffness of the material. To validate the analysis, a test system has been developed to measure both the damping and the stiffness properties of the material under axial excitation. The system has been shown to be capable of measuring frequency dependent damping coefficients of less than 1%. Tests on the wavy composite have led to a characterization of the material over a broad range of frequency and temperature that agrees well with predicted values, and allows one to design the wavy composite lay-up to give a desired performance at a given temperature and frequency. Damping measurements as high as 30% have been measured, while maintaining stiffness properties that are comparable to materials such as aluminum and titanium, using standard graphite fibers.

11:00

2aSA5. Forced response condensations applied to the eigenvalue analysis of damped structures. Zhao Xianfeng and J. Gregory McDaniel (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

The eigenvalue analysis of nonproportionally damped structures has received much recent attention, as most of the algorithms to solve undamped problems do not directly extend to the damped case. In many cases, one is only interested in finding the eigenvalues that lie in a region of the complex plane. In this presentation, we propose an algorithm that exploits this fact. The algorithm begins by finding the eigenvalues of the undamped problem and tracks the eigenvalues as damping is gradually added. For each increment of damping, new eigenvalue estimates are obtained by forcing the structure at the previous eigenpair and using the forced response vectors to condense the matrices. In this way, one may be assured of finding all of the eigenvalues in a vertical strip of the complex plane. Features of the algorithm will be discussed and initial proof-of-concept results will be presented. [This material is based upon work supported by NSF under Grants Nos. 9978747 and 9984994.]

11:15

2aSA6. A new type of vibration damper based on flexural wave propagation in laminated wedges of power-law profile. Victor V. Krylov (Dept. of Aeronautical and Automotive Eng., Loughborough Univ., Loughborough, Leicestershire LE11 3TU, UK, V.V.Krylov@lboro.ac.uk)

Flexural waves propagating towards sharp edges of elastic wedges are slowing down and growing in amplitudes as they approach the edges. Especially interesting phenomena may take place in wedges having cross sections described by a power-law relationship between their local thickness and the distance from the edge, with m being a power-law exponent. In particular, the theory shows that, for m equal or larger than 2—in free wedges, and 5/3—in immersed wedges, the incident flexural waves can become trapped near the edge and never reflect back, i.e., the above structures represent acoustic ''black holes'' for flexural waves. Because of the deviations of real manufactured wedges from the ideal power-law profiles, first of all because of ever-present edge truncations, the corresponding reflection coefficients are always far from zero. Therefore, in reality such homogeneous wedges can not be used as effective vibration dampers. In the present paper it is demonstrated that the situation can be significantly improved by depositing an absorbing thin layer on the wedge surfaces. In this case the reflection coefficients can be drastically reduced even for wedges with noticeable truncations. Thus, the combination of the power-law profile and a thin absorbing layer can result in a new effective method of damping flexural vibrations.

Session 2aSC

Speech Communication: Segmental Acoustics, Perception and Production (Poster Session)

Betty H. Tuller, Chair
Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, Florida 33431

Contributed Papers

All posters will be on display from 8:30 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 10:15 a.m. and contributors of even-numbered papers will be at their posters from 10:15 a.m. to 12:00 noon.

2aSC1. An acoustical study on Lai vowel length. Kenneth Vanbik and Rungpat Roengpitya (Dept. of Linguist, 1203 Dwinelle Hall, Univ. of California at Berkeley, Berkeley, CA 94720)

Lai (Chin), a language from the Sino-Tibetan language family, has a contrastive vowel length. In this paper, the relationship between the vowel quantity and quality in Lai and the relative duration of these vowels and the following consonant were studied. The speech of an adult male native speaker of Lai, uttering in a frame sentence words having the shape /pVC/ where V was long and short versions of the five phonemic vowel qualities and C was either /t/ or /n/, was digitized. The $F1$, $F2$, and the duration of the vowels were measured as well as the duration of the final nasal. The data revealed that the long:short vowel duration ratio was about 2.5 for the tokens with a final stop and about 2.86 for the tokens with a final nasal. Vowels preceding a stop tended to be overall shorter than vowels preceding a nasal; final nasals after short vowels were significantly longer than after long vowels; short vowels were more centralized than long vowels. Thus although the vowels described as long:short are longer versus shorter, there are other correlated differences serving this contrast.

2aSC2. Acoustic characteristics of fricatives in Arabic. Mohamed Al-Khairy and Ratree Wayland (Prog. in Linguist., 4131 Turlington Hall, P.O. Box 115454, Univ. of Florida, Gainesville, FL 32611-5454)

Previous work on English fricatives suggests that a conglomerate of acoustic cues exist to differentiate among fricatives produced at different places of articulation [e.g., Jongman (2000)]. These include spectral peak location, spectral moments, $F2$ onset frequency, noise amplitude, noise duration, etc. The goal of this study is to investigate whether or not acoustic characteristics served to classify English fricatives place of articulation can be used to classify fricatives place of articulation in Makkan Arabic. Unlike English, this dialect of Arabic includes velar, pharyngeal, and glottal fricatives in its phonological inventory. Thus of great interest is whether or not cues to fricatives place of articulation thus far discovered will be able to successfully differentiate among these rare fricatives. All acoustic measurements found in Jongman's 2000 work including locus equations will be used to explore this important question, and results of this study will be compared and contrasted with those of earlier studies.

2aSC3. Effects of consonant aspiration on Mandarin tones. Ching X. Xu and Yi Xu (Dept. of Commun. Sci. and Disord., Northwestern Univ., Evanston, IL 60208, xxq@northwestern.edu)

In this paper we present an experiment on the influence of consonant aspiration in Mandarin. A minimal pair, /da/ and /ta/, which differ only in aspiration, are produced in carrier sentences by native speakers of Mandarin. It is found that consonant aspiration affects the fundamental frequency ($F0$) of the following vowel. The onset $F0$ of a tone is higher following unaspirated consonants than following aspirated consonants, and the magnitude of difference is related to both the current and preceding tones. However, the later part of a tone tends to be lower following an unaspirated consonant than an aspirated one. These results seem to favor the speech aerodynamic model proposed by Ohala (1976), which predicts a smaller value of airflow at the onset of the vowel following an aspirated stop than an unaspirated one. The current results, however, seem to provide further details about the mechanism of the aerodynamic effect. Additionally, results of detailed analyses seem to suggest a possible source of the existing controversies over the exact effect of aspiration on $F0$. The discrepancy in the previously reported effects may be partially due to the difference in how the first few vocal cycles of the vowel are measured in different studies.

2aSC4. Tonal assimilation in Thai. Rungpat Roengpitya and John Ohala (Dept. of Linguist., Univ. of California, Berkeley, CA 94720)

Thai has five distinctive tones: mid, low, rising-falling, high, and falling-rising. In this paper, the behavior of Thai tones when they were concatenated with other tones was studied. The target utterances were Thai meaningful and nonsense words with the shape C1VC2, where the C1 was a nasal, the V was /i/ or /i:/, bearing each of the five tones, and C2 was either a nasal or a voiceless stop. The targeted words were spoken by three native speakers of Thai in isolation and in five different tonal contexts, digitally recorded, and measured for the fundamental frequency of the targeted words and the following words. The results showed that the tones of the targeted words normally started at the vowel onset but their offset carried over to the beginnings of the tones of the following words. This carryover effect [previously noted by Xu (1997) for Mandarin tones] occurred more for the targeted words with the final nasals than for the ones with the final stops.

2aSC5. Variation of reaction times for the recognition of voiced plosives with preceding noise duration. William Ainsworth (MacKay Inst. of Commun. and Neurosci., Keele Univ., Staffs ST5 5BG, UK)

Voiced plosives in noise which precedes the speech are recognized correctly more often than in noise which begins at the start of the speech. This effect may be due to noise adaptation of firing thresholds of neurones in the cochlear nerve effectively performing spectral subtraction, or it may be due to perceptual streaming at a higher level of the auditory system. In order to investigate this further, reaction times for recognition have been measured. It was found that the reaction time initially reduces for a preceding noise of 100–200 ms, then increases as the noise duration is lengthened (at least up to 800 ms). Furthermore it was found that the reaction time is reduced when the recognition score is higher. Assuming that recognition consists of processing which increases with the noise

duration but decreases for good internal matches, a model of reaction time as a function of preceding noise duration has been developed. It consists of two negative exponential functions: one a function of noise duration and the other a function of recognition score. Good agreement between experimental data and the model has been found. [Work supported by EU TMR Contact No. FMRX-CT97-0150.]

2aSC6. Phonological and acoustic-phonetic cues in recognizing words ending in nonrelease stop consonants. Janielle Lugo-Marin and Cynthia M. Connine (Dept. of Psych., State Univ. of New York, Binghamton, NY 13902)

This study investigates the effects of probabilistic likelihood and of certain acoustic–phonetic cues in the recognition of word-final nonrelease stop consonant stimuli. Recognition of the word was examined by assessing the activation of an embedded word (inconsistent with a release) relative to the activation of the whole word. The missing release was presented in contexts that predict a higher (a following consonant) or lower (a following vowel) likelihood of a release. The contribution of acoustic–phonetic cues was examined by including word-final stop consonants that were voiced (do not require release for recognition) or voiceless (release required for recognition). Four experiments were designed to examine the effects of the probabilistic likelihood for the release and of acoustic–phonetic cues in the activation of a canonical version of a word, when the final stop-consonant release is not present. Priming of the embedded word was greater for the voiceless relative to the voiced stimuli, and also for the following consonant context relative to vowel context. In the absence of the following context, the embedded word was primed for the voiceless but not for the voiced stimuli, and the whole word was primed to a greater extent for the voiced relative to the voiceless stimuli.

2aSC7. Relative timing of oral and laryngeal motions in voiceless consonant production: A comparison of single and geminate consonants in Japanese. Akinori Fujino (NTT Commun. Science Labs., NTT Corp., 3-1 Morinosato Wakamiya, Atsugi-shi, Kanagawa 243-0198, Japan, fujino@idea.brl.ntt.co.jp), Tokihiko Kaburagi (Dept. of Acoust. Design, Kyushu Inst. of Design, Minami-ku, Fukuoka 815-8540, Japan), Masaaki Honda (NTT Commun. Sci. Labs., NTT Corp., Atsugi-shi, Kanagawa 243-0198, Japan), Emi Z. Murano (Dept. of Otolaryngol., Univ. of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan), and Seiji Niimi (Dept. of Speech-Lang. Pathol. and Audiol., Intl. Univ. of Health and Welfare, Otawara-shi, Tochigi 324-0011, Japan)

It is known that temporal coordination of oral and laryngeal articulations is important in voiceless consonant production. In this study, the relative timing of oral constriction and glottal opening is investigated for single and geminate voiceless consonants in Japanese. The motions of the jaw, lips, tongue, and larynx of three male speakers were examined using EMA (electromagnetic articulograph), PGG (photoglottograph), and EGG (electroglottograph) techniques. A set of /iie eCe desu/ and /iie eQCe desu/ sentences, containing voiceless consonants /s, t, ϕ, p/, was used for the test utterances. /QC/ in the sentence shows a geminate form in Japanese. The experiment data showed that the interval between the peak of glottal opening and oral release onset for geminates was longer than for singles and that consonant duration was also longer. For each form, the longer the consonant duration was, the longer the interval became. These temporal relations were found for all speakers, although the ratio of the two intervals was different for each speaker. The results suggest that, for both single and geminate consonants, articulators are controlled based on the consonant duration dependence of the oral-glottal timing. [Work supported by CREST, JST.]

2aSC8. Transconsonantal coarticulatory patterns in VCV utterances: Effects of a bite block. Yana Yunusova and Gary Weismer (Waisman Ctr. and Dept. of Communicative Disord., Univ. of Wisconsin, Madison, Madison, WI 53705-2280)

The concept of coordinative structures, or articulatory synergies, envisions a collection of articulators organized to achieve a specific articulatorly or acoustic goal. The mandible figures prominently in concepts of articulatory synergies because of its potential interaction with labial and lingual shaping of the vocal tract. What happens to these hypothesized synergies when one component is taken out of the collective? Bite block articulation is a common experimental approach to eliminating the jaw from its synergistic role in articulation, and studies have shown that the speech mechanism is able to reorganize its target configurations almost immediately when speaking with a fixed jaw. In the current study we examine vowel-to-vowel, transconsonantal effects with and without a bite block. Although there is one study [Sussman, Fruchter, and Cable, J. Acoust. Soc. Am. (1995)] showing stability of coarticulatory effects in biteblock conditions, that conclusion was based primarily on locus equations. In the current study, we hypothesize that more traditional acoustic measures of right-to-left and left-to-right coarticulation will show that reducing an articulatory synergy by holding one of its components constant will result in different-from-typical coarticulatory behaviors. [Work supported by NIDCD Award No. DC 000319.]

2aSC9. A cross-linguistic study of articulatory timing in liquids. Bryan Gick, Fiona Campbell, and Sunyoung Oh (Dept. of Linguist., Univ. of British Columbia, E270-1866 Main Mall, Vancouver, BC V6T 1Z1, Canada, gick@interchange.ubc.ca)

A number of recent studies have observed that the timing of articulatory gestures is systematically different in prevocalic versus postvocalic allophones of English liquids. In English, the more anterior (consonant-like) gestures tend to occur at the peripheries of syllables, while more posterior (vowel-like) gestures occur nearer the syllable peak. These facts have allowed for a greatly simplified account of phonological processes involving English liquids, such as ''vocalization'' of postvocalic liquids, r- or l-coloring of vowels, and ''dark'' and ''light'' allophones of /l/. However, no such studies have been made of timing in liquids of other languages. In this study, ultrasound data were collected from one to three subjects for each of English /l/ and /r/, Quebecois French /l/, Mandarin /r/, and Korean /l/. Differences in intergestural timing were measured for prevocalic versus postvocalic allophones. Results show similar timing effects to those previously observed for English, with posterior gestures occurring earlier in postvocalic allophones. This property is consistent with phonological processes, such as pre-liquid laxing, observable in these languages. Quebecois /l/ and Mandarin /r/ were also interesting in exhibiting no measurable dorsal gesture at all in their initial allophones. [Research supported by the NSERC.]

2aSC10. MRI motion imaging of articulatory movement-coarticulation effects in velar consonants. Noriko Suzuki (School of Dentistry, Showa Univ., 2-1-1 Kitasenzoku, Ohta-ku, Tokyo, Japan), Tatsuya Kitamura (Faculty of Information, Shizuoka Univ., 3-5-1 Johoku, Hamamatsu, Shizuoka, Japan), Shinobu Masaki (ATR Human Information Processing Res. Labs.), and Ken-ichi Michi (School of Dentistry, Showa Univ., 2-1-1 Kitasenzoku, Ohta-ku, Tokyo, Japan)

Magnetic resonance imaging (MRI) techniques have been used in speech research for measuring the vocal tract. Recently, MRI motion imaging has been developed and applied to the study of speech production. In this study, the coarticulatory effects of Japanese velar stop consonants were measured using a synchronized sampling method. The MRI technique employed is a synchronized sampling technique which is a modification of one originally developed for cardiac motion imaging. Speech samples consisted of /aka/ and /aki/, including the Japanese velar stop consonant /k/ and vowel /i/, in normal subjects. Results depicted motion of the speech organs in the midsagittal plane. When velar consonants were

followed by a back vowel /a/, the back of the tongue contacted the posterior part of the soft palate, and when followed by front vowel /i/, tongue contact was more toward the anterior of the soft palate. Thus the tongue contact position for Japanese velar stop consonants was found to be strongly influenced by the following vowel.

2aSC11. Phase transitions in a repetitive speech task as gestural recomposition. Kenneth J. de Jong, Byung-jin Lim, and Kyoko Nagao (Dept. of Linguist., Indiana Univ., 322 Memorial Hall, Bloomington, IN 47405, kdejong@indiana.edu)

Tuller and Kelso (JSHR, 1991) show the relative timing of peak glottal opening and oral closure in repetitive /ip/ production differs between /ip/ and /pi/ productions. Also, as repetition rate increases, the timing in /ip/ merges with that in /pi/. The authors interpret this to indicate two stable regions in glottal-to-oral phasing, one for codas and one for onsets. The current paper examines a similar corpus of glottal transillumination traces of repeated syllables, which includes both labial and lingual stops. Preliminary examination of peak glottal timing with respect to acoustic landmarks indicating closure shows the same pattern of timing differences at slow rates and neutralization at fast rates. However, peaks in the transillumination traces often do not indicate glottal abduction for the consonant, but rather indicate a relatively open glottis for vocalic voicing, happening in the context of glottal stoppage at the onset of the vowel. In addition, coronal codas exhibit cycle-to-cycle variation from glottalized to abducted voiceless forms. Thus while rate does induce stepwise changes in glottal activity, phasing, in the sense of relative timing, does not account for many of the rate effects. Rather, rate appears to be inducing a recomposition of the gestures which comprise the speech.

2aSC12. Gestural timing effects in the "perfect memory" sequence observed under three rates by electromagnetometry. Mark K. Tiede (M.I.T.-R.L.E., Cambridge, MA and Haskins Labs, 270 Crown St., New Haven, CT 06520, tiede@haskins.yale.edu), Joseph Perkell, Majid Zandipour (M.I.T.-R.L.E., Cambridge, MA), and Melanie Matthies (Hearing Res. Ctr., Boston Univ. and M.I.T.-R.L.E., Cambridge, MA)

In a well-known example due to Browman and Goldstein [Lab. Phon. I, 341–376 (1990)] the /ktm/ sequence in the phrase ''perfect memory'' is contrasted between careful (list) and fluent production conditions. X-ray microbeam data were used to show that although in the fluent case coproduction of the /m/ can mask the acoustic releases of the /k/ and /t/, both stops are nonetheless articulated. The current work uses EMMA data to examine this sequence in greater detail: 8 subjects produced the phrase in a carrier context under normal, fast, and clear rate conditions. Results confirm that tongue dorsum and tip movements toward velar and apical closure occur regardless of rate and observable acoustic effect. In addition, while movement amplitudes decreased somewhat as rate increased, little variation in the durations associated with the consonant gestures was observed. Instead, changes in rate primarily effected V to V duration and the relative phasing of the velar and apical closing gestures: the tongue tip maximum (/t/) lagged that of the tongue dorsum (/k/) in clear speech, was aligned with it in normal speech, and generally anticipated it in the fast rate condition. [Work supported by NIH.]

2aSC13. Correlations between vowel and the subsegments of an obstruent in CV and VCV sequences. Hansang Park (Dept. of Linguist., Calhoun Hall, Univ. of Texas, Austin, TX 78712)

It is well known that there is a significant negative correlation between a consonant and the preceding or following vowel. However, it has seldom been specified what relationship lies between the subsegments (closure, frication, and aspiration) of an obstruent and the preceding or following vowel in Korean. This work investigates the relationship between the vowel and the subsegments of a Korean coronal obstruent in CV and VCV sequences. There was a significant negative correlation between the gap and vowel ($r=-0.900$) in utterance initial CV sequence but no significant correlation between frication and vowel. On the other hand, there was a significant negative correlation between the preceding vowel and closure ($r=-0.510$) in the VCV sequence but no significant correlation between the preceding vowel and other subsegments. There was a significant negative correlation between the following vowel and gap ($r=-0.545$) but no significant correlation between the following vowel and other subsegments. Generally, there was a significant correlation between the vowel and an immediately adjacent segment than between the vowel and remote segments. Correlation was stronger in the CV sequence than in the VCV sequence for the same pair of variables. Speech Communication Best Student Paper Award.

2aSC14. Dynamic center-of-gravity effects in consonant–vowel transitions. Ewa Jacewicz, Nandini Iyer, Robert A. Fox, and Lawrence L. Feth (The Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210)

Lublinskaja has reported that dynamic changes in the spectral center-of-gravity (COG) of one formant in selected Russian vowels led to changes in the identification of the vowel. Movement of the COG was effected by simultaneous amplitude modulation of two formants placed at the end points of the desired frequency transition [Iyer *et al.*, J. Acoust. Soc. Am. **109**, 2294 (2001)]. The present study explores whether such COG effects will extend into the processing of consonant–vowel transitions (CVs). A [dɑ]–[gɑ] continuum was generated using a Klatt synthesizer with a 50-ms CV transition and a 200-ms vowel. The onset of *F*3 varied from 2018–2818 Hz in 80-Hz step. The *F*3 of the vowel remained constant at 2527 Hz. Three different stimulus sets differed in terms of whether the *F*3 transition was a: (1) formant, (2) FM tone, or (3) virtual frequency (VF) glide. In the VF stimuli, the frequencies of two tones matched the onset or offset frequencies of the *F*3 transition and modulation of their amplitudes produced a gliding spectral COG. Preliminary results indicate that listener performance is not strongly affected by changing the means by which spectral changes to *F*3 are made. [Research supported by a grant from OSU's College of Social and Behavioral Sciences to L. Feth and an INRS award from NIH to R. Fox.]

2aSC15. Perceptual and articulatory noise in vowel imitation. Gautam K. Vallabha and Betty Tuller (Ctr. for Complex Systems and Brain Sci., Florida Atlantic Univ., 777 Glades Rd., Boca Raton, FL 33431)

Speakers exhibit systematic errors when asked to imitate self-produced vowels [B. Tuller and G. Vallabha, J. Acoust. Soc. Am. **107**, 2855 (2000)]. These results were replicated (3 speakers, 45 systematically distributed self-produced vowel targets, 10 imitations per target) and compared to a simulation of the effects of articulatory and perceptual noise. Articulatory noise was modeled using the Haskins Articulatory Synthesizer. The tongue body center (TBC) was interpolated between the configurations for /i/, /a/, and /u/ and 90 of the intermediate configurations were perturbed by adding independent Gaussian noise (sd=1 mm) to the *x* and *y* coordinates of the TBC. Perceptual noise was modeled by (1) transforming each of the 90 intermediate vowels to *F*1/*F*2 bark (''perceptual'') space and (2) adding independent Gaussian noise (sd=vowel discrimination limen=0.28 barks) to each bark coordinate. The models did not match the directions of the imitation errors made by the speakers. However, the distribution of the error magnitudes, relative to the mean imitation, was approximately Rayleigh (0.28) across both the models and the speakers. These results indicate that while speech perception and/or production may have intrinsic Gaussian noise, this noise is not responsible for the systematicity of the imitation errors. [Work supported by NIMH.]

2aSC16. Vowel formant discrimination in quiet, speech-shaped noise, and babble for natural speech. Chang Liu and Diane Kewley-Port (Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47405)

Vowel formant discrimination in quiet and in noise for naturally spoken vowels was investigated. The stimuli were seven isolated vowels /i, u, I, ɛ, ʌ, æ, a/ recorded from a female talker and resynthesized using STRAIGHT with and without formant shifts. Vowels were presented in two noise types, long-term speech-shaped noise (LTSSN) and twelve-talker babble. ΔF was measured for $F1$ and $F2$ in quiet and noise with five signal to noise ratios (SNRs), $+4$, $+2$, 0, -2, and -4 dB. Adaptive tracking procedures were used, interleaving quiet and SNR levels for one formant at a time. Overall, results suggested that formant frequency, SNR, and noise type had significant effects on ΔF. For either LTSSN or babble, ΔF as a function of formant frequency showed four significantly different patterns, depending on the SNRs. ΔF as a function of SNR also showed several different patterns for different formants. A surprising result was that ΔFs were higher at -2 dB than those at -4 dB in babble, particularly in the $F2$ region. In addition, performance in babble was significantly better than that in LTSSN at comparable SNRs. These results will be examined by partial loudness models. [Work supported by NIHDCD-02229.]

2aSC17. Effects of synthesis fidelity on vowel identification: Role of spectral change and voicing source. Peter F. Assmann and William F. Katz (School of Human Development and Callier Ctr. for Community Disord., The Univ. of Texas at Dallas, Box 830688, Richardson, TX 75083)

Recent studies have shown that synthesized versions of American English vowels are more accurately identified when natural time-varying changes in the formant frequencies are preserved rather than flattened. A limitation of these experiments is that vowels generated with cascade formant synthesis are generally less accurately identified than natural vowels. To overcome this limitation, a high-quality analysis–synthesis system was used to reexamine the effects of spectral change. Using this new technique, synthesized versions of 12 American English vowels spoken by adults and children were identified as accurately as natural vowels. Two experiments confirmed the beneficial effects of preserving the time-varying changes in the formants, both in vowels synthesized with pulsed excitation (as in voiced speech) and noise excitation (as in whispered speech), and verified that identification accuracy does not decline when the fundamental frequency is held constant. However, in contrast to earlier findings, (i) whispered vowels were identified as accurately as the voiced versions, and (ii) the benefits of time-varying spectral change were greater than previously found with cascade formant synthesis. The findings are consistent with recent studies showing that the measured effects of spectral and temporal manipulations in vowels can vary as a function of synthesis fidelity.

2aSC18. Recognizing vowels from narrow band spectra. James M. Hillenbrand (Speech Pathol. and Audiol., Western Michigan Univ., Kalamazoo, MI 49008) and Robert A. Houde (RIT Res. Corp., Rochester, NY 14623)

Most models of vowel recognition begin with a representation consisting of either a smooth envelope spectrum or formant frequencies extracted from those envelope spectra. In the present report we describe a classification method that performs pattern matching by comparing high-resolution vowel spectra (i.e., spectra that retain harmonics of the voice source) with a set of smoothed spectral-shape templates for each of the vowels derived by averaging harmonic spectra from a large database of /hVd/ utterances spoken by men, women, and children [Hillenbrand *et al.*, J. Acoust. Soc. Am. **97**, 3099–3111 (1995)]. [Work supported by NIH.]

2aSC19. Intelligibility of narrow-band speech is enhanced by unmodulated flanking noise: Restoring intelligibility by masking of silence. Peter W. Lenz, James A. Bashford, Jr., and Richard M. Warren (Dept. of Psych., Univ. of Wisconsin-Milwaukee, Milwaukee, WI 53201, plenz@uwm.edu)

The intelligibility of notch-filtered sentences may improve by 10% or more when a large spectral gap is filled with noise rather than silence [Warren *et al.*, P&P **59**, 275–283 (1997)]. The present study shows that a comparable ''spectral completion'' (SC) effect occurs when a narrow band of speech is surrounded by complementary noise. In Experiment 1, intelligibility of 2/3-octave everyday sentences, centered at 1.5 kHz [2000-order FIR filtering, slopes greater than 1000 dB/octave], was 65% without noise, 72% with complementary low-pass or high-pass noise, and 76% with both noise bands present (noise spectrum level at 20 dB relative to the speech). Experiment 2 found a comparable SC effect with partially masked speech: The 2/3-octave sentences were about 40% intelligible when masked by narrow-band noise having the same spectral limits (relative spectrum level again 20 dB), but were 50% intelligible when the noise was broadband. Hence, out-of-band noise enables listeners to make better use of information within narrow-band speech, perhaps by masking an otherwise misleading absence of energy in normally coactive regions of the spectrum. Under partial masking conditions, this SC effect can produce an apparent release from masking in the absence of comodulation. [Work supported by NIH.]

2aSC20. Dual band speech intelligibility: Implications of some novel observations concerning amplitude mismatch and asynchrony. Richard M. Warren, James A. Bashford, Jr., and Peter W. Lenz (Dept. of Psych., Univ. of Wisconsin-Milwaukee, Milwaukee, WI 53201, rmwarren@uwm.edu)

Two 1/3-octave bands of everyday sentences (center frequencies 1 and 3 kHz) were employed. To avoid the major contribution of slopes (even when 100 dB/octave) upon intelligibility [Warren and Bashford, J. Acoust. Soc. Am. **106**, L47 (1999); Warren *et al.*, J. Acoust. Soc. Am. **108**, 1264 (2000)], nearly vertical slopes were employed (4000 order FIR filtering producing 2000 dB/octave slopes). Heard alone at 75 dB, intelligibilities were 5% (1 kHz) and 10% (3 kHz); together, the synergistic score was 80%. When the 3-kHz band was kept at 75 dB and the 1-kHz band's amplitude was decreased systematically, intelligibility remained unchanged (1.4%) from 75 to 45 dB (amplitude ratio of 1000:1). But when both bands were at their normal relative levels, and one band was delayed relative to the other, intelligibility dropped to half with asynchrony of only about 35 ms (approximately half the duration of the average phoneme), decreasing to an intelligibility score of a single band when asynchrony approximated phonemic duration. It is suggested that broadband speech processing relies upon cross-frequency comparison of dynamic changes in local patterns with little regard to differences in pattern amplitude. This comparison appears to be based upon units approximating the duration of the average phoneme. [Work supported by NIH.]

2aSC21. Phonemes at the cocktail party: Insights on the processing of competing speech signals. James A. Bashford, Jr., Richard M. Warren, and Peter W. Lenz (Dept. of Psych., Univ. of Wisconsin-Milwaukee, Milwaukee, WI 53201, bashford@uwm.edu)

A recorded word repeated over and over is heard to undergo a series of illusory changes (verbal transformations) to other syllables and words in the listeners lexicon. When cocktail party listening conditions are simulated by adding a second image of the repeating word through dichotic presentation with an interaural delay preventing fusion, the resulting two distinct lateralized images of the word undergo independent transformations [Lenz *et al.*, J. Acoust. Soc. Am. **107**, 2857 (2000)]. The present study found that adding a second image of the same word increased the extent of phonetic distortions for each image, whereas presenting a different competing word reduced the transformation rate by one-half when the

competitor differed by one phoneme (e.g., ''dark'' paired with contralateral ''bark'') and by three-fourths when it differed in all phonemes (e.g., ''dark'' versus contralateral ''light'' or ''might''). These interactions were independent of semantic relations and occurred with competitors from both same and different speakers. In contrast, transformations were unaffected by a nonspeech competitor (a repeating contralateral door slam). These and additional results indicate that shared resources at an abstract level of phonological analysis provide the basis for the independent syllabic/lexical organizations derived for competing speech signals. [Work supported by NIH.]

TUESDAY MORNING, 4 DECEMBER 2001 ROOMS 122/123, 8:55 TO 11:00 A.M.

Session 2aSP

Signal Processing in Acoustics, Acoustical Oceanography, Underwater Acoustics, Biomedical Ultrasound/ Bioresponse to Vibration: Acoustic Image Reconstruction Using Tomographic Techniques

Sean K. Lehman, Cochair
Lawrence Livermore Laboratory, P.O. Box 808, L-154, Livermore, California 94551

Michael Silevitch, Cochair
Northeastern University, 360 Huntington Avenue, Room 235FR, Boston, Massachusetts 02115

Chair's Introduction—8:55

Invited Papers

9:00

2aSP1. The Center for Subsurface Sensing and Imaging Systems (CenSSIS): Addressing the challenge of implementing a strategy of "diverse problems similar solutions." Michael Silevitch (Northeastern Univ., CenSSIS, 360 Huntington Ave., Boston, MA 02115, msilevit@lynx.neu.edu)

CenSSIS has recently been designated by the National Science Foundation as an Engineering Research center (ERC). The CenSSIS mission is to unify the methods and tools used to detect and image biomedical and environmental–civil objects or conditions that are underground, underwater, or embedded in the human body. In the area of physics based signal processing and image understanding, methods of multiview tomography play a key role. Beginning with a brief description of CenSSIS, this talk will then focus upon a discussion of a unified framework that tries to articulate a methodology for characterizing end to end subsurface sensing and imaging problems. Several examples will be used to illustrate the ideas, including ones dealing with acoustic tomography.

9:25

2aSP2. Acoustic wave diffraction tomography. Anthony J. Devaney (Northeastern Univ., ECE, 360 Huntington Ave., Boston, MA 02115, tonydev2@aol.com)

The classical theory of Diffraction Tomography (DT) is formulated within the context of the acoustic and elastic wave equations. A number of applications for the generalized theory are reviewed including ultrasound tomography in variable density and elastic wave media and cross-well geophysical tomography and offset vertical seismic profiling (VSP). The general theory of DT is compared and contrasted with classical focus-on-transmit and focus-on-receive confocal acoustic imaging as well as with various standard *ad-hoc* imaging methods often employed in applications. The use of slant stacking for converting point source to plane wave insonification is reviewed as is the theory of acoustic tomography in known layered background media. A number of computer simulations and results from experimental data are presented.

9:50

2aSP3. Focusing by spatiotemporal inverse filter: Application to heterogeneous and absorbing medium. Michael Tanter, J.-F. Aubry, J.-L. Thomas, and M. Fink (Laboratoire Ondes et Acoustique, ESPCI, 10 rue Vauquelin, 75005 Paris, France, michael.tanter@espci.fr)

In most applications of ultrasound, a first step consists in focusing a broadband ultrasonic beam deeply inside the medium. Nevertheless, sound speed, density, or absorption heterogeneities may strongly degrade the focusing. Different techniques have been developed to correct such aberrations (time reversal, speckle brightness, . . .). Recently, we investigated a new broadband focusing technique, the spatiotemporal inverse filter, based on the inversion of the propagation operator relating a transducer array to a set of control points inside the medium. As this inversion is achieved both in space and time, it allows one to calculate the set of temporal signals to be emitted by each array element for an optimal focusing on a chosen control point. Potential applications have been experimentally investigated in heterogeneous and absorbing media where classical geometrical focusing is highly degraded. The strong improvement achieved by inverse filter compared to time reversal and classical focusing emphasizes the great potential of multichannel systems having the ability to apply completely different waveforms on each element. Particularly, intraplate echoes suppression in NDT, high quality focusing through a human skull, as well as hyper-resolution in reverberating media highlight its great interest in various fields such as medical imaging, nondestructive testing, room acoustics, and underwater acoustics.

10:15–10:30 Break

Contributed Papers

10:30

2aSP4. Resolution enhancement algorithm utilizing multiple reflections for image reconstruction in the conduit. Kenbu Teramoto and Kohsuke Tsuruta (Dept. of Mech. Eng., Saga Univ., Saga-shi 840-8502, Japan)

In the acoustical endscopy, due to the physical limitations, the transducer array is composed of a small number of elements and each interspacing is larger than the acoustical wave length that is called a sparse array system. In such case, the volumetric image reconstruction process becomes ill-posed caused by the interference of the reverberation in the tube. Consequently, the reconstructed image is interfered by widely spread strong spurious lobes. This paper proposes a novel volumetric image reconstruction algorithm for a 3D sparse aperture acoustical holography, which visualizes the forward-looking image in the conduit tube. The proposed algorithm utilizes prior knowledge about both field, object, and observed signal as (1) the support of the object is limited within the conduit, and (2) the multiple reflected wave field scattered by the object is statistically independent of the reverberation in the tube. The proposed sparse array acoustical imaging system utilizes, therefore, the multiple reflected wavefronts in order to enhance the resolution. By experiments, it is proven that the proposed algorithm has the following improvements: (1) Targets can be identified when multiple reflections exist in the tubes. (2) Artifacts caused by the spurious lobes can be reduced.

10:45

2aSP5. Two-dimensional synthetic focusing using large apertures—Deconvolution of the spatial impulse response. Fredrik Lingvall and Tadeusz Stepinski (Signals and Systems Group, Dept. of Mater. Sci., Uppsala Univ., P.O. Box 528, SE-751 20 Uppsala, Sweden, fredrik.lingvall@signal.uu.se)

The synthetic aperture focusing technique (SAFT) has been used successfully in both medical and nondestructive evaluation (NDE) applications of ultrasound. Transducer size is an important issue when implementing SAFT algorithms. The classical SAFT method is based on the assumption that a point source is used for emitting ultrasonic waves, which means in practice that the transducer aperture is so small compared to the wavelength that it emits spherical waves. If the transducer aperture is large its spatial impulse response (SIR) significantly differs from a Dirac pulse and nonspherical waves will be generated. Therefore, to successfully perform synthetic focusing using large apertures, the transducer's SIR must be taken into account. A new method for compensating transducer SIR in synthetic focusing is presented in the paper. The method employs a stochastic, time-domain Wiener filtering technique for two-dimensional synthetic focusing. Results of the experiments performed in water using a linear phased array are presented to demonstrate the performance of the proposed method. The use of an array enabled altering the aperture without changing the electrical characteristics of the whole system. The results obtained using the proposed technique for finite apertures are compared to those obtained with a classical SAFT algorithm.

TUESDAY MORNING, 4 DECEMBER 2001 ROOM 304, 8:25 A.M. TO 12:00 NOON

Session 2aUW

Underwater Acoustics: Seabed Characterization and Environmental Inversion

Tracianne B. Neilsen, Chair
Applied Research Laboratories, University of Texas at Austin, P.O. Box 8209, Austin, Texas 78713-8029

Chair's Introduction—8:25

Contributed Papers

8:30

2aUW1. Benchmark workshop for geoacoustic inversion techniques in range dependent waveguides. Ross Chapman (School of Earth and Ocean Sci., Univ. of Victoria, P.O. Box 3055, Victoria, BC V8W 3P6, Canada, chapman@uvic.ca), Stan Chin-Bing, David King (Naval Res. Lab., Stennis Space Ctr., MS 39529-5004), and Richard Evans (SAIC, New London, CT)

This paper summarizes the results from the ONR/SPAWAR Geoacoustic Inversion Techniques (IT) Workshop held in May 2001. The format of the workshop was a blind test to estimate unknown geoacoustic profiles by inversion of synthetic acoustic field data for vertical and horizontal array geometries in range dependent shallow water waveguides. The fields were calculated using COUPLE/RAM for three range-dependent test cases: a monotonic slope; a shelf break; and an intrusion in the sediment. Geoacoustic profiles were generated to simulate sand, silt and mud sediment environments. Different approaches for inverting the field data were presented at the workshop: model-based methods based on normal modes, parabolic equation or ray theory; perturbative methods; methods using transmission loss (TL) data; methods using vertical or horizontal array data. The geoacoustic profiles inverted by the different methods are compared using a metric based on calculation of transmission loss using the estimated profiles for geometries and source frequencies that were not used in the inversion. The results demonstrate the effectiveness of present day inversion techniques, and indicate the limits of their capabilities in range dependent environments. [Work supported by ONR.]

8:45

2aUW2. Geoacoustic inversion techniques workshop: Test cases 4 and 5. James K. Fulford, David B. King, and Stanley A. Chin-Bing (Naval Res. Lab., Acoust. Div., Stennis Space Center, MS 39529)

Acoustic transmission loss and reverberation data from two exercises were selected as inversion test cases for the Geoacoustic Inversion Techniques workshop, May 2001. The data was collected using air deployed sonobuoys with standard 0.82 kg SUS charges as sources. The acoustic data was presented in octave bands for frequencies from 25 to 800 Hz, and a 1/3 octave band for 3.5 and 6.5 kHz center frequencies. Environmental data consisted of bathymetry, and sound speeds gathered by AXBT's. Background geophysical information was not given to the participants. Inversion results were encouraging. Selected examples will be shown. [Work supported by ONR/NRL and SPAWAR.]

9:00

2aUW3. Inversion for source and environmental parameters using a broadband rotated coordinates technique. Tracianne B. Neilsen and David P. Knobles (Appl. Res. Lab., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

The problem of localizing and tracking a source in the shallow ocean is often complicated by uncertainty in the environmental parameters. Rotated coordinates [M. D. Collins and L. Fishman, J. Acoust. Soc. Am. **98**, 1637–1644 (1995)] calculated for a broadband cost function are used in simulated annealing to invert for both source and environmental parameters. The resulting broadband rotated coordinates identify the coupling and relative sensitivity the parameters over the specified bounds. Multiple sets of broadband rotated coordinates, each corresponding to a different set of bounds, are used to systematically decouple the parameters. The cost function minimized in the inversion algorithm is model-data cross-phone spectra summed coherently over phone pairs, frequency, and time sequence. The results of applying this inversion method to simulated array data are presented along with preliminary results using data from the Area Characterization Test II experiment performed on the New Jersey continental shelf. [Work supported by ONR.]

9:15

2aUW4. Matched-field inversion of seabed geoacoustic properties complemented by chirp sonar surveys. Altan Turgut and Stephen Wolf (Naval Res. Lab., Washington, DC 20375)

Matched-field inversion of range-dependent seabed parameters is studied using broadband (525–725 Hz) acoustic data collected on the New Jersey Shelf during the SWARM95 experiment. An efficient global optimization scheme is used to minimize an objective function defined by the Bartlett processor output in both beam and mode space. With the input from chirp sonar bathymetry and sediment thickness data, a broadband acoustic field is calculated by a coupled normal mode model and range-dependent geoacoustic properties are inverted. An independent chirp sonar inversion of acoustic impedance profiles was also performed along the same track by using chirp sonar data. The inversion results indicate good agreement between two inversion methods and show their effectiveness if they are used as complementary to each other. The effect of a strongly range-dependent water column on the inversion performance and possible broadband inversion using single hydrophone is also studied. [Work supported by ONR.]

9:30

2aUW5. Using an adjoint model to invert for unknown perturbations to a presumed environment. Paul Hursky, Michael B. Porter (Sci. Applications Intl. Corp., 1299 Prospect Ave., Rm. 305, La Jolla, CA 92037), Bruce Cornuelle, William S. Hodgkiss, and William A. Kuperman (Scripps Inst. of Oceanogr., Univ. of California, San Diego, La Jolla, CA 92093-0701)

Adjoint modeling is an inversion technique that has the potential to sharply reduce the number of modeling runs typically needed to achieve an inversion. Although adjoint models are being used in oceanographic and meteorological modeling, they seem to have been overlooked by the acoustics community. A single adjoint modeling iteration calculates an adjustment to forward model control inputs (initial conditions and boundary conditions) needed to correct for a mismatch between predicted and measured system parameters. We will present how the adjoint of an acoustic propagation model can be used to trace a mismatch in predicted and measured acoustic parameters back to the perturbations in the environmental parameters needed to correct for the mismatch.

9:45

2aUW6. Seabed characterization from statistical analysis of high-frequency data. Samantha J. Dugelay, Richard J. Brothers, and Gary J. Heald (QinetiQ Bincleaves, Newton's Rd., Weymouth, Dorset DT4 8UR, UK)

Results are presented from statistical analysis of experimental and simulated high-frequency seabed backscatter data. The statistical laws employed in the analysis were the Rayleigh distribution, the Rayleigh mixture, and the K-law. The significance of the distribution parameters and their estimation from the data are explained. The link between the parameters and seabed type has been theoretically examined through a detailed study of the variation of statistical laws with grazing angle. Experimental data were collected using a 150-kHz wide-band sonar at three different sites, providing different sediment types. The validity of the above statistical laws was investigated and found to be dependent on sonar geometry. The estimated parameters of the three distributions have been examined for their potential in seabed characterization. Simulations have been compared with experimental data for validation purposes. These results have then been used to investigate the theory developed beyond the scope of the recorded data.

10:00–10:15 Break

10:15

2aUW7. Seabed layer discrimination using normal incidence backscatter returns in the 1–10 kHz frequency band. Gary J. Heald (QinetiQ, Winfrith Technol. Ctr., Winfrith Newburgh, Dorchester, Dorset DT2 8XJ, UK) and Paul C. Hines (Defence Res. Establishment Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada)

It has been established that the first and second seabed returns from a monostatic echo sounder can be used as a seabed classification tool. These systems operate on the principle that the first echo is purely monostatic, whereas the second echo can be treated as a bistatic return with a virtual source located at a height of 2–3 times the water depth. Typically these systems operate in the range of 100 kHz so that bottom roughness is the dominant scattering mechanism. In the present work we apply this procedure to lower frequencies. This should provide additional discrimination information since contributions from the seabed volume become significant. Additionally, it provides information about the nature of seabed layers that may be present. However, this comes at the cost of further complicating the model. In this paper we modify the high frequency model to be applicable at lower frequencies and examine the effect of sub-seabed layering.

10:30

2aUW8. Acoustic scattering from stratified marine sediments. Anatoliy N. Ivakin (APL-UW, 1013 NE 40th St., Seattle, WA 98105, on leave from Andreev Acoust. Inst., Svernika 4, Moscow 117036, Russia, anivakin@yahoo.com)

There are various mechanisms of seabed scattering. Normally, they are attributed to two main types of bottom irregularities: roughness of the sediment interfaces and volume fluctuations of the acoustical parameters. However, there is ambiguity in the description of these irregularities. It is known, e.g., that roughness can be exactly expressed in terms of volume perturbations near the flat interface. On the other hand, there is a class of volume perturbations within continuously stratified media, which can be described in terms of the roughness of isosurfaces of acoustical parameters. Here, all these cases are considered using a unified approach to

volume and roughness scattering [A. Ivakin, J. Acoust. Soc. Am. **103** (1998)]. This approach unifies also the cases of continuous and discrete stratification (layering). Rough interfaces between layers are replaced by continuous transition layers having rough isosurfaces of the acoustical parameters. First-order perturbations of these parameters are proportional to their vertical gradients and to displacement (roughness) of the isosurfaces within the transition layers. The model gives the volume scattering strength as a function of the wave thickness of the transition layers. In the particular case of low frequencies, where the transition layers are thin in comparison with the wavelength, the model provides an exact transition to existing results for scattering from rough interfaces and, thus, naturally links volume and roughness mechanisms of seabed scattering at low and high frequencies within a unified approach.

10:45

2aUW9. Extraction of modal backscattering matrix from reverberation data in shallow-water waveguide. Part I. Theory. E. C. Shang (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093), T. F. Gao (Inst. of Acoust., Chinese Acad. of Sci., Beijing, PROC), and D. J. Tang (APL, Univ. of Washington, Seattle, WA 98105)

Extracting the bottom backscattering information from reverberation data in shallow-water waveguide is an attractive but difficult issue. In previous works, some *a priori* assumption (for instance, the Lambert's Law, or the separability of the backscattering matrix) has been made in order to solve this problem. In this paper, new approaches are proposed. The modal backscattering matrix can be extracted directly from reverberation data without any *a priori* assumption on scattering coefficient. The backscattering information inversion is based on the following three components: (1) mode-filtering at the receiving vertical array, (2) changing the point source depth to obtain different incident mode excitation, and (3) using mode stripping in the waveguide as an additional mode filter. For lower frequency the linear inversion is presented, and for higher frequency the sequential inversion with angle interval smoothing is proposed. [Work supported by ONR.]

11:00

2aUW10. Seabed and sub-bottom classification using measurements of normal incidence backscatter from 1–10 kHz. John C. Osler, Paul C. Hines, and Darcy J. MacDougald (Defence Res. Establishment Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada)

It has been established that the first and second normal incidence acoustic returned from the seabed can be used as a classification tool to discriminate between different sediments. Typically this is accomplished using monostatic echosounder systems that operate in the range of several 10 s of kHz so that bottom roughness at the seabed interface is the dominant scattering mechanism. Model results (Hines and Heald, Proceedings of the Institute of Acoustics Conference on Acoustical Oceanography, Southampton, UK, April 2001) indicate that at lower frequencies, normal incidence returns from the sub-bottom should provide an additional discrimination tool without significantly corrupting the information contained in the interface scatter returns. This paper presents an experimental methodology for making these measurements in the 1 to 10 kHz frequency band using a vertical line array of receivers and a downward-looking superdirective projector array. Results from experiments at the ONR Strataform area and Scotian Shelf will be compared with independent geo-acoustic information.

11:15

2aUW11. Validity of backscatter strength models at high (110–170 kHz) frequencies. Sarah J. Page, Richard J. Brothers, and Patricia A. G. Thomson (QinetiQ Bincleaves, Newton's Rd., Weymouth, Dorset DT4 8UR, UK)

Sonar performance models use several low-level acoustic models to predict the effect of the environment on reverberation and noise. Many of these acoustic models, particularly those predicting seabed scattering strength, were developed at lower frequencies than those currently used by both the scientific and mine-hunting communities. This paper examines the popular Applied Physics Laboratory of the University of Washington (APL-UW) acoustic backscatter strength model and tests its validity in the extended frequency band of 110–170 kHz. Acoustic and environmental data (gathered as part of a joint program with NATO SACLANTCEN) from several sites are collated and compared in this paper. This provides compelling evidence regarding the validity of the APL-UW backscatter model at these higher frequencies.

11:30

2aUW12. Comparison of bistatic scatter modeling with experimental tank data at 250 kHz. Richard J. Brothers, Gary J. Heald (QinetiQ Bincleaves, Newton's Rd., Weymouth, Dorset DT4 8UR, UK), Paul McCloghrie, Philippe Blondel, and Nicholas G. Pace (Univ. of Bath, Bath BA2 7AY, UK)

Experimental data gathered in the laboratory at the University of Bath have been compared to bistatic scattering models. The data were gathered at 250 kHz over different controlled sediment types (sand/gravel/cobble), and at a range of scattering angles using a directional source and omnidirectional receiver. Stringent system calibration was undertaken, along with extensive pre-experimental analysis, in order to guarantee high-fidelity scattering data within the confines of a tank environment. Having established the validity of the scattering models, they have been incorporated into a bistatic sonar model. This is achieved by integrating a function that includes the scattering cross-section at each point on the sediment surface, as well as spreading and absorption losses. This model is designed to investigate optimum sonar parameters for the bistatic detection of objects on and in the seabed.

11:45

2aUW13. Acoustic field of hydroacoustic antenna arranged near a rough surface. Tran. Huy Dat (Inst. of Hydromechanics of Natl. Acad. of Sci. of Ukraine, Zhelyabov Str. 8/4, Kiev 03035, Ukraine)

The characteristics of an averaged field of a hydroacoustic antenna arranged near a penetrable ocean surface with large-scale random roughness is formulated. The model of an antenna is described as an activity cylindrical surface, restricted absolutely rigid toroid. The given model rather convincingly mirrors actual design features of modern hydroacoustical antennas. The averaged field of a hydroacoustic antenna is analyzed depending on a degree of roughness of boundary, sizes of antennas, field pattern on activity surfaces, way of the view of space and bottom penetration. It showed that, the rough surface could result into differences in distribution of averaged values of a far field of hydroacoustic antennas up to 20% in comparison with a case of free environments, the value bottom penetration, after which one of the distribution values of the mean characteristics of a near field of the antenna practically coincides appropriate values for a free environment, is established. By the given calculation and discussion, it is followed that, the ocean rough surface could be considered as the natural factor resulting in decreasing of acoustic shadow. Nevertheless, bistatic calculations of coherent field suggest useful strategies for improving in design a new model of the hydroacoustic antenna.

Session 2pAAa

Architectural Acoustics: Speech Privacy Design in Office Spaces

Kenneth P. Roy, Chair
Innovation Center, Armstrong World Industries, 2500 Columbia Avenue, Lancaster, Pennsylvania 17603

Chair's Introduction—1:00

Invited Papers

1:05

2pAAa1. Speech privacy design and masking sound systems. Kenneth P. Roy (Armstrong World Industries, Innovation Ctr., 2500 Columbia Ave., Lancaster, PA 17403)

The term ''speech privacy'' when applied to the office building segment must be considered both from the point of view of the person talking and from that of the listener. The achievement of adequate speech privacy for a particular situation is based on both the architectural design of the space, and the acoustical factors including speech and noise characteristic. The signal to noise ratio is the key to privacy performance, and this will be addressed by review of a case study for a recent office application. Electronic masking sound is a key concept in the attainment of speech privacy; and a new approach to masking system design will be reviewed, including an improved speaker concept, source DSP, and modeling tools.

1:25

2pAAa2. Speech privacy in closed offices: Comparison of languages and accent. Yong Ma, Daryl J. Caswell (Dept. of Mech. & Manufacturing Eng., Univ. of Calgary, Calgary, AB T2N 1N4, Canada), Jim T. Goodchild (Smed Intl., Calgary, AB T2C 4T5, Canada), and Liming Dai (Univ. of Regina, SK S4S 0A2, Canada)

Up to now, most objective assessments for speech privacy and speech intelligibility, such as articulation index (AI), speech intelligibility index (SII), early-to-late sound ratio (Clarity), and speech transmission index (STI), are evaluated by subjective measurements primarily based on studies that incorporate only the English language. In today's multicultural environment, it is necessary to study the impact on speech privacy of different languages and accents. In this work, subjective measurements were conducted in closed offices by using English, Mandarin Chinese (a tonal language), and accented English. Both close-set and open-set test materials, including single words, sentences, and articles were employed in the measurement. The results revealed the evaluation differences in speech privacy between two languages, as well as between native English and accented English. The subjective measurement results were also compared with the objective measurement indices AI, STI, and Clarity.

1:45

2pAAa3. Speech privacy comparison using Articulation Index (AI) and Speech Privacy Noise Isolation Class (NIC) in an open office. Heakyung Yoon and Vivian Loftness (School of Architecture, Carnegie Mellon Univ., MM Hall Rm 410, Pittsburgh, PA 15213, hcy@andrew.cmu.edu)

Two speech privacy measurements, Articulation Index (AI) standardized by the American National Standard Institute in 1969 and Speech Privacy Noise Isolation Class (NIC, Noise Isolation Class Prime) adapted from the Noise Isolation Class (NIC) by U. S. General Services Administration (GSA) in 1979, have been claimed as objective tools to measure speech privacy in open offices. To evaluate which of them is a better tool in terms of speech privacy in a dynamic open office environment, the measurements were taken in the field. AIs and NICs in the current workplace settings as well as in the modifications of partition heights and workplace configurations have been measured following ASTM E1130 (Standard Test Method for Objective Measurement of Speech Privacy in Open Offices Using Articulation Index) and GSA test PBS-C.2, respectively.

2:05

2pAAa4. Noisy offices for simple tasks and quiet offices for complex tasks? Heakyung Yoon and Vivian Loftness (School of Architecture, Carnegie Mellon Univ., MM Hall Rm 410, Pittsburgh, PA 15213, hcy@andrew.cmu.edu)

Many current office plans are designed as open environments to increase knowledge sharing, which is supposed to increase performance through more informal interactions between workers. However, according to workplace surveys, poor acoustics in open office environments have been reported to decrease workers performances. The literature regarding performance with quiet, ambient, and masking sound in offices show that the kinds of performed tasks in organizations or groups may be the main factor to decide whether open or closed office environments are appropriate. People engaged in simple office tasks perform better with masking sounds

than with quiet or ambient office sound conditions [Loewen *et al.*, ''Cognitive and arousal effects of masking office noise,'' Environment and Behavior **24**, 381–395 (1992)]. However, when people need to use serial-recall functions in their complex tasks, quiet conditions may be better than other sound environments in offices [Banbury *et al.*, ''Disruption of office-related tasks by speech and office noise,'' Br. J. Psychol. 499–517 (1998)].

Contributed Papers

2:25

2pAAa5. Speech levels in open plan offices. Alf Warnock and Wing T. Chu (Inst. for Res. in Construction, Natl. Res. Council, Canada)

A key variable for predicting speech privacy in open offices is the level of speech used. Since the environment influences vocal effort, speech levels needed to be measured in typical open office situations to more accurately calculate speech privacy in such offices. This paper describes measurements of voice levels used in face-to-face conversation in nine open offices in the Ottawa/Hull Area. Speech from 118 subjects was recorded and analyzed. Average voice levels in each office ranged from 48 to 55 dBA. The overall mean was 51 dBA with a standard deviation of 3.4 dBA. Levels from the current study were significantly lower than those in ASTM standard E1130, but the overall average and spectrum agree well with Pearsons' results for casual conversation (K. S. Pearsons, R. L. Bennett, and S. Fidel, ''Speech levels in various noise environments,'' EPA-6001-77-025, U. S. Environmental Protection Agency, May 1977). The paper describes the procedures and the results obtained. This work was done as part of a project for Public Works and Government Services Canada.

2:40

2pAAa6. Directivity of human talkers. Wing T. Chu, Alf Warnock (Inst. for Res. in Construction, Natl. Res. Council, Canada), and Jean-Charles Guy (INSA de Lyon, France)

As well as the level of speech, the directivity of the sound field surrounding the human head is important when estimating speech privacy in open offices. Information on this subject was found to be sparse so the sound field surrounding human talkers was recorded by two sets of eight microphones fixed on two orthogonal meridian arcs in an anechoic chamber. Talkers were positioned at the center of the two arcs of the microphone arrays and rotated through six positions at 15° increments to survey the whole sound field. Twenty male and twenty female subjects were used. No significant differences in directivity were detected between male and female talkers although the mean spectra were different. There was good agreement between the field and laboratory measurements of the male voice spectra though some differences were found for the female voice spectra. No significant differences in directivity or frequency content were detected between English and French talkers. Similar directivities were obtained for the average human talker and the Bruel and Kjaer Head and Torso Simulator. Agreement with other published information is fairly good. This work was done as part of a project for Public Works and Government Services Canada.

TUESDAY AFTERNOON, 4 DECEMBER 2001 ROOM 315, 3:30 TO 4:30 P.M.

Session 2pAAb

Architectural Acoustics: The Technical Committee on Architectural Acoustics Vern O. Knudsen Distinguished Lecture

David Lubman, Chair
14301 Middletown Lane, Westminster, California 92683

Chair's Introduction—3:30

Invited Paper

3:35

2pAAb1. Acoustic concerns related to multi cultural societies. Anders Chr. Gade (Oersted, DTU, Section of Acoust. Technol., Tech. Univ. of Denmark, Bldg. 352, DK-2800 Lyngby, Denmark, acg@oersted.dtu.dk)

Immigration has increased cultural diversity in western societies. The process of integrating immigrants into their host countries can be smoothed if acousticians learn to recognize (1) the acoustic traditions of immigrant cultures and (2) the specific acoustic needs of the new society members. Two related projects are discussed. The ''Cahrisma'' project (Conservation of Acoustical Heritage by the Revival and Identification of the Sinan's Mosque Acoustics) is sponsored by the European Commission and carried out in cooperation among researchers in Turkey, Malta, Italy, France, Switzerland, and Denmark. Its purpose is to combine visual and acoustical concerns in the identification, conservation, and restoration of architectural heritage. It focuses on the famous Turkish mosques of the Osmannic architect Sinan. Some of the acoustic features of these large domed buildings and of muslim liturgy will be presented. Another project was conducted by students at the Danish Technical University, who found that speech intelligibility for non-native listeners is much more sensitive to poor speech conditions than for native listeners. Also discussed are the unique features of culturally based urban soundscapes. It is suggested that these soundscapes can provide comfort to recent immigrants by increasing their sense of being ''at home.''

Session 2pAB

Animal Bioacoustics: Topics in Animal Bioacoustics

David A. Mann, Chair
College of Marine Science, University of South Florida, 140 Seventh Avenue, South, Saint Petersburg, Florida 33701

Contributed Papers

1:30

2pAB1. A comparison of visual and acoustic marine mammal monitoring methods. Adam S. Frankel and Kathleen Vigness-Raposa (Marine Acoust., Inc., 901 N. Stuart St., Ste. 708, Arlington, VA 22203)

Concern about the potential effect of increased oceanic noise on marine mammals has led to the consideration of a variety of at-sea marine mammal monitoring methods. A recent marine mammal survey utilized both visual observation and passive acoustic monitoring. Visual observations were conducted using 7×50 binoculars during the day and generation III night vision devices (NVDs) at night. Acoustic data were collected with a towed hydrophone array and analyzed with a customized, PC-based acoustic workstation. Seventy-seven visual sightings were made. Forty sightings were made at night and 37 were made in daylight. The effective range of the NVDs was estimated through observations of a target at known distances; 50% detection rating was achieved at 130 m. There were 98 acoustic detections. Of the 77 visual sightings, 42 were also detected acoustically. Six sightings were made without comparable acoustic detection, and 19 sightings were made when there was no acoustic monitoring. Conversely, 56 acoustic detections were made without visual sightings. These data suggest that, for species that commonly vocalize, a combined acoustic/visual survey will increase detection probabilities. Furthermore, night vision devices have the potential to be an effective observation tool for marine mammals. [Work sponsored by ONR.]

1:45

2pAB2. Sonic boom penetration under a wavy air–water interface. H. K. Cheng (Univ. of Southern California, Los Angeles, CA 90089), Adam Fincham (Laboratoire des Ecoulements Géophysiques et Industriels, CNRS-UJF-INPG, BP 53, Grenoble 38041, France), Tony Maxworthy (Univ. of Southern California, Los Angeles, CA 90089), and John R. Edwards (Environ. Management Div., Space and Missle Systems Ctr., Los Angeles AFB, CA 90245)

The interaction of a sonic boom with a wavy air–water interface is investigated theoretically and experimentally to ascertain its effect on sonic boom noise penetration in deep water. The Mach number is assumed to be less than the water-to-air sound speed ratio of 4.53. Unlike the problem with a flat (nonwavy) interface, the interaction with a surface-wave train can produce *everfescent* (nonattenuating) wave components, in addition to *evanescent* wave components underwater. The overall effect results in an attenuation rate with a depth which is much lower than predicted by the flat-ocean (Sawyers) model; it overwhelms the flat-ocean wavefield in deep water, even though it has a secondary effect at the surface. The expectation is substantiated by both analysis and laboratory measurements. The latter were performed in a water channel with wave-makers and projectile over flights. Detailed measurements confirm both the wave packet characteristics of the deep water waveform and the frequency downshift property predicted by the theory. Underwater sound-pressure level, frequency range, and waveform characteristics from sonic booms of aircraft and space-launch operations are examined with the wavy-ocean model.

2:00

2pAB3. Using high-frequency acoustics to protect marine mammals. Joseph S. Johnson (Chief of Naval Operations, N7747, 2000 Navy Pentagon, Washington, DC 20350-2000), William T. Ellison (Marine Acoust., Inc., Litchfield, CT 06759), Peter Stein (Sci. Solutions, Inc., Nashua, NH 03063), and Clayton H. Spikes (Marine Acoust., Inc., Arlington, VA 22203)

A joint industry-government research effort has designed, developed, and tested a high-frequency marine mammal monitoring (HF/M3) system to ensure that marine mammals are not exposed to high levels of low-frequency sonar signals. The system's nominal operating characteristics are 30–40 kHz frequency, a variable bandwidth of 1.5–6 kHz, 3%–4% duty cycle, 10–40 m pulse length, 3–4 s pulse repetition rate, and 220 dB (*re*: 1 μPascal at 1 m) source level. HF/M3 operating parameters are based on *in situ* acoustic modeling and low-power probe-pulse signals, which define the local ocean acoustic environment. Operating protocols include source level adjustments to ensure approaching marine mammals or sea turtles are unaffected by HF/M3 signals. System capability has been tested and evaluated in 7 sea trials. Estimated probabilities of detection for a 20-m stationary whale at 160 m depth are from 98% at 1-km range to 72% at 2-km range. The ability to track animals via multiple detections virtually eliminates randomly distributed false alarms. Single-ping probabilities of detection suggest that any marine mammal (and possibly sea turtle) within 1 km can be expected to be detected multiple times—approximately once every 50 s.

2:15

2pAB4. Interactions of Florida manatees (*Trichechus manatus latirostris*) with simulated fishing gear and a pinger: Implications for preventing entanglements. A. E. Bowles, C. Alves, and R. A. Anderson (Hubbs-Sea World Res. Inst., 2595 Ingraham St., San Diego, CA 92109)

Nine stranded, rehabilitated and captive-born Florida manatees (*Trichechus manatus latirostris*) were exposed to simulated crab-pots and gillnetting to quantify interactions (70 trials). A pinger (Dukane–Seacom Netcom 1000, 300 ms ping at 4 s intervals, level ~130 dB *re* 1 μPa from 10 to 80 kHz) was presented with and without the crab pot during 7 of the trials. Manatees were observed for 30 min (baseline), then exposed to the object for 30 min. During exposures, they manipulated objects extensively, with little or no evidence of neophobic avoidance. Most contacts (70%) were made with the mouth; the remainder involved the pectoral flippers, body, and paddle. Entanglements sufficient to require human intervention occurred in four trials. Manatees spent 36% more time in the test area when objects were present than during baseline trials, indicating a positive motivation to manipulate. The pinger stimulated brief startle reactions in two of four female manatees but in none of five males. Crab pots with pingers were contacted as much as the simulated crab pot alone (51% vs 41% by females). This motivation to manipulate suggests that manatees cannot be protected from fishing gear by pingers.

2:30–2:45 Break

2:45

2pAB5. Hyperbolic location errors due to insufficient numbers of receivers. John L. Spiesberger (Dept. of Earth and Environ. Sci., Univ. of Pennsylvania, 240 S. 33rd St., Philadelphia, PA 19104-6316)

Animal locations are sometimes estimated by measuring the difference in travel times of their sounds at pairs of receivers. Ideally, each difference specifies the animal's location to a hyperboloid, and sufficient numbers of intersecting hyperboloids specify the location. Most bioacoustic publications state that one needs three receivers to locate animals in two spatial dimensions and four receivers in three spatial dimensions. However the actual numbers of receivers required for locations in two and three dimensions are four and five, respectively. Significant location errors can result from using insufficient numbers of receivers. Methods are provided for determining locations of ambiguous source-location regions.

3:00

2pAB6. Songlike vocalizations and infrasound from the Sumatran rhinoceros. Elizabeth von Muggenthaler (Fauna Commun. Res. Inst., P.O. Box 1126, Hillsborough, NC 27278), Paul Reinhart (Cincinnati Zoo, 3400 Vine St., Cincinnati, OH 45220), and Brad Lympany (Asheboro, NC)

Within the last ten years the Sumatran rhino (*Dicermoceros sumatrensis*) population has dropped 50%, and only 200–300 individuals are left in the world. The oldest living species of rhino in evolutionary terms, Sumatran rhinos are solitary, although males and females are seen together during courtship. Their native habitat is dense tropical forest and mountain moss forest. They are the smallest living rhino, standing 0.09–1.5 m tall, and are covered in course, reddish-brown hair. Three Sumatran rhinos, housed at the Cincinnati Zoo, were recorded from 1–3 m. Two Statham Radio microphones, and two Sony TCD-D8 DAT recorders recorded from 9 Hz to 22 KHz. Analysis, including FFTs, spectrographs, and filtering, were performed using National Instrument's Polynesia. The rhinos proved to be extremely vocal, producing signals almost constantly. Distinct calls, including several types of ''eeps,'' 70 Hz–4 kHz (57–92 dB); ''whales,'' 100 Hz–3.2 kHz (87 dB); and ''whistle-blows,'' 17 Hz–8 kHz (100 dB) were discovered. The ''whistle-blow'' has high dB infrasound that would be advantageous for use in the rhino's forest habitat. Some Sumatran rhino vocalizations sound similar to and resemble (under analysis) some humpback whale signals.

3:15

2pAB7. The felid purr: A healing mechanism? Elizabeth von Muggenthaler (Fauna Commun. Res. Inst., P.O. Box 1126, Hillsborough, NC 27278, L@animalvoice.com)

A current hypothesis suggests the purr indicates contentment, however, cats purr when they are severely injured or frightened. Forty-four felids were recorded including cheetahs, ocelots, pumas, domestic cats, and servals. A Sony TCD-D8 Digital Audio Recorder (DAT) and Statham Radio microphones recorded the purrs. FFTs and spectrographs were performed using National Instrument's Polynesia. An accelerometer was also used to measure domestic cat purrs. Every felid in the study generated strong frequencies between 25 and 150 Hz. Purr frequencies correspond to vibrational/electrical frequencies used in treatment for bone growth/fractures, pain, edema, muscle growth/strain, joint flexibility, dyspnea, and wounds. Domestic cats, servals, ocelots, and pumas produce fundamental, dominant, or strong frequencies at exactly 25 Hz and 50 Hz, the two low frequencies that best promote bone growth/fracture healing [Chen *et al.*, Zhong. Wai Ke Za Zhi. **32**, 217–219 (1994)]. These four species have a strong harmonic exactly at, or within 2 Hz of 100 Hz, a frequency used therapeutically for pain, edema, wounds, and dyspnea. An internal healing mechanism would be advantageous, increasing recovery time and keeping muscles and bone strong when sedentary. [Published with permission from the *New Zealand Veterinary Journal*; work supported by Endevco.]

TUESDAY AFTERNOON, 4 DECEMBER 2001 ROOM 305, 1:00 TO 5:00 P.M.

Session 2pAO

Acoustical Oceanography and Physical Acoustics: Turbulence and Finestructure Studies I: Surface and Bottom Boundary Layer Turbulence

James F. Lynch, Chair

Woods Hole Oceanographic Institution, 203 Bigelow Building, Woods Hole, Massachusetts 02543

Invited Papers

1:00

2pAO1. Dynamics of bottom boundary layers in the coastal ocean. John Trowbridge (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543, jtrowbridge@whoi.edu)

The bottom boundary layer is the region adjacent to the sea floor, with a thickness typically on the order of meters to tens of meters, where turbulence generated by bottom drag produces vertical mixing of mass, heat, and momentum. Oceanic boundary layers are influenced not only by turbulent mixing, but also by planetary rotation, stratification, topography, surface waves, internal waves, and interaction with the erodible sea floor. Historically, understanding of the oceanic bottom boundary layer has been based on classical results in engineering and meteorology. However, recent theory and measurements have revealed unique features of oceanic boundary layers. Outstanding problems include measuring the interaction of the flow with the erodible sea floor; measuring the spatial scales of turbulent motions that accomplish vertical transport of mass, heat, and momentum; and understanding the processes that control the intensity and scale of boundary layer turbulence, particularly in stable stratification.

2:00

2pAO2. Laboratory and field measurements of breaking-induced turbulence: Implications for ocean acoustics. W. Kendall Melville (Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0213)

The entrainment of air by breaking surface waves, its breakup into bubbles, and their transport in the surface layers by breaking-induced turbulence and Langmuir circulations have important consequences for active and passive acoustical oceanography. Recent laboratory and field work related to these issues will be presented. The description of breaking-induced turbulence and air entrainment by just field experiments is very difficult and there are some advantages to separating the issues into those that must be addressed in the field and those that are better studied in the laboratory. For example, the statistics of breaking cannot be resolved by laboratory measurements, whereas the detailed structure of the turbulence is probably better studied in the laboratory. New results from airborne measurements of breaking statistics, and laboratory (and field) measurements of breaking-induced turbulence will be presented and their implications for upper ocean acoustics will be discussed. Similarities between breaking-induced turbulence and Langmuir circulations will be discussed. [Work supported by ONR and NSF.]

2:25

2pAO3. 2D imaging of breaking waves with a pulse-coherent phased array. Jerome Smith (Scripps Inst. of Oceanogr., 0213 UCSD, La Jolla, CA 92093-0213)

Breaking waves represent a serious challenge in the study of the air–sea interface. Entrainment of air and the expulsion of droplets and particles by breaking waves are of primary importance to material fluxes in particular and air/sea exchanges in general. Observations of motion near the surface is key to understanding the kinematics and dynamics of this elusive interface. Unfortunately, it is nearly impossible to obtain measurements closer than a meter or two below actively breaking wave crests in the open ocean. Here, measurements from an up-looking phased-array sonar system are presented. The system provides measurements over pies about 22 degrees wide by 16 m maximum range, with resolution 1.5 degrees by 5 cm. The instrument was 13.5 m below the mean surface, so the measurement cell size near the surface is about 5 cm (vertical) by 30 cm (along-wind). Sequences of intensity images can be analyzed for motion and relative scatterer (bubble) density. Data presented were obtained as a wave of about 2.5 m height crest to trough passed through the field of view while breaking (one of several breakers captured over a two-week deployment from R/P FLIP).

2:50

2pAO4. Doppler measurement of velocity structure beneath surface gravity waves. David Farmer, Johannes Gemmrich, and Svein Vagle (Inst. of Ocean Sci., Sidney, BC V8L 4B2, Canada)

The acoustic Doppler method has proved particularly useful for the investigation of small scale velocity structure associated with wave motions, shear instability, and turbulence. The incoherent bistatic approach permits measurement of the velocity field at meter scales within the crests of large and steep waves, allowing direct comparison with predictions in the region where theory is most likely to break down. At the other end of the scale, coherent Doppler measurement at megahertz frequencies can be used to identify and estimate the magnitude of turbulence beneath breaking waves. We contrast these different approaches using observations with a bistatic incoherent Doppler sonar placed on the floor of the North Sea and an array of monostatic coherent sonars deployed on the ocean surface off the California coast. The observations illustrate the power of Doppler techniques in resolving various aspects of wave dynamics which have been inaccessible using traditional methods.

3:15

2pAO5. Characteristics of turbulence in the surf zone. John Trowbridge (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543, jtrowbridge@whoi.edu) and Steve Elgar (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Near-bottom measurements from an array of acoustic Doppler velocimeters over a sand beach produce direct estimates of turbulent shear stress, by means of a differencing procedure that rejects contributions from surface waves, and indirect estimates of dissipation, by means of an inertial-range model. Measurements of winds, waves, and currents provide a context for analysis of the turbulence measurements. The analysis focuses on an alongshore momentum equation, semiempirical relationships between turbulent shear stress and alongshore velocity, and simplified budgets for energetics. The records are dominated by three events when the instruments were in the surf zone. Near-bottom turbulent shear stress was well correlated with, but smaller than, the sum of wind and wave forcing. Vertical gradient of alongshore velocity was larger than indicated by the Prandtl–Karman law of the wall. Drag coefficient was smaller during breaking waves than during unbroken waves. Shear production of turbulent kinetic energy was approximately equal to dissipation, and both were two orders of magnitude smaller than the depth-averaged rate at which the shoaling wave field lost energy to breaking, indicating that breaking-induced turbulence did not penetrate to the seabed.

3:40–3:55 Break

3:55

2pAO6. Coherent Doppler profiler measurements of near-bed suspended sediment fluxes. Carolyn E. Smyth and Alex E. Hay (Dalhousie Univ., 1355 Oxford St., Halifax, NS B3H 4J1, Canada, csmyth@phys.ocean.dal.ca)

Vertical profiles of vertical turbulence intensity and vertical sediment fluxes were collected by an acoustic Coherent Doppler Profiler during two experiments: at Queensland Beach, NS and SandyDuck97. Measurements are investigated as a function of bedstate from low energy vortex ripples to high energy flat bed. Ensemble-average vertical turbulence intensity profiles exhibit a peak within the wave boundary layer at heights of O (1 cm) above bottom for all bedstates. Peak average turbulence intensities are relatively independent of bedstate, varying by no more than 50% despite a factor of 7–10 variation in average wave energy. This otherwise remarkable observation can be understood from the corresponding decrease in the physical roughness of the bed, associated with the different observed bedstates. Estimated wave friction factors are highest for low energy rippled beds and smallest for flat bed conditions. Estimates of the vertical suspended sediment flux partitioned into mean, wave, and turbulent components, show that in general, there is a balance between upward fluxes and downward settling; except immediately above the bed, and except for the case of a stationary ripple field. The suspended sediment flux coherence indicates enhancement at incident wave frequencies, with the largest coherence for flat bed conditions very near the bed.

4:10

2pAO7. Measurements of turbulence in the wave boundary layer using high frequency acoustics. Eric Terrill, Ronan Gray, and Ken Melville (Scripps Insitution of Oceanogr., Marine Physical Lab., Mail Code 0213, La Jolla, CA 92093-0213)

The properties of the turbulent boundary layer at the air–sea interface exhibits departures from classical wall bounded shear flows as a result of the free surface characterized by surface waves. As surface waves grow in the presence of wind, they become unstable and break, which leads to the injection of turbulence and bubbles into the water column. While an understanding of these physical processes is critical to fields of oceanography and mid-/high-frequency ocean acoustics, there remains much to be learned due in part to the difficulty of obtaining turbulence measurements at the air–sea interface as well as signal processing issues associated with separating turbulent water motions from wave orbital motions. One maturing solution to this measurement problem is the use of $O(1)$ MHz pulse to pulse coherent Doppler sonars that allow direct wave number measurements in the inertial subrange of the turbulence. Recent field measurements of near-surface turbulence have been conducted by our group with this sonar technology both offshore Pt. Conception California and Hawaii. Measurement results will be discussed, including observations regarding the relationship between the intermittancy of the turbulent dissipation to the overlying surface wave field. [Work supported by ONR and NSF.]

4:25

2pAO8. Observations of turbulence in ocean wave boundary layers. Timothy P. Stanton (Oceanogr. Dept., NPGS, Monterey, CA 93943)

A bistatic geometry, coherently sampled, acoustic Doppler profiler has been developed to measure turbulent stresses in wave-forced ocean boundary layers. The profiler is typically configured to measure over-resolved profiles of three-component velocity vectors sampled in $O(2\ \mathrm{cm}^3)$ sample volumes at 20 Hz over a 60 cm range above the ocean bed. Examples of bottom boundary layer observations made at Duck, NC and Monterey, CA during two field experiments illustrate approaches to measuring Reynolds stresses and shear in thin oscillatory boundary layers over a rippled bed. During the ONR-funded Shoaling Waves Experiment (SHOWEX) the profiler was mounted on an instrumented frame at 12 m depth and continuously measured velocity profiles and changes in wave-induced bed morphology through a succession of storms over a two month period. In the second example, observations of mean currents, wave-induced velocities, and turbulence were made in the surf zone just offshore from a steep-sloping beach.

4:40–5:00

Panel Discussion

TUESDAY AFTERNOON, 4 DECEMBER 2001 ROOM 223, 2:00 TO 5:30 P.M.

Session 2pBB

Biomedical Ultrasound/Bioresponse to Vibration: Topical Meeting on Physics of Ultrasound in Relation to the Biology of its Therapeutic Effects: Ultrasound with Adjunctive Agents

Lawrence A. Crum, Chair

Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

Chair's Introduction—2:00

Invited Papers

2:05

2pBB1. A possible physical mechanism of ultrasound-activated gene delivery: Shear stress generated by microstreaming. Junru Wu (Dept. of Phys., Univ. of Vermont, Burlington, VT 05405, jwu@zoo.uvm.edu)

Gene therapy provides potentially effective treatment of many diseases. Sonoporation and ultrasound induced transfection can promote the uptake and expression of foreign DNA by a cell. *In vitro* studies [Bao *et al.*, Ultrasound Med. Biol. **23**, 953–959 (1997); Greenleaf *et al.*, Ultrasound Med. Biol. **24**, 587–595 (1998); Ward *et al.*, J. Acoust. Soc. Am. **105**, 2951–2957 (1999); Ultrasound Med. Biol. **26**, 1169–1175 (2000); Guzman *et al.*, J. Acoust. Soc. Am. **110**, 588–596 (2001)] have shown that the presence of contrast

agents dramatically enhances the efficiency of the transfection and also reduces the required acoustic pressure. Recent experimental studies and numerical calculations also demonstrated that the shear stress associated with microstreaming surrounding encapsulated bubbles of contrast agents near living cells driven by 0.1 MPa acoustic pressure amplitude ultrasound a 1 MHz or 2 MHz may be large enough to generate reparable sonoporation to the cells. This presentation will review above-mentioned publications and discuss the possible physical mechanism involved.

2:25

2pBB2. Ultrasound contrast agents to enhance and monitor therapy. Katherine Ferrara, James Chomas, Donovan May, John Allen (Dept. of Biomed. Eng., Univ. of California Davis, Davis, CA 95616), Rachel Pollard, and Erik Wisner (Univ. of California Davis, Davis, CA 95616)

New techniques for the assessment of tissue perfusion and the local delivery of drugs have been developed using ultrasound contrast agents. A perfusion estimator, based on the destruction and wash-in of ultrasound contrast agents, has been developed and applied to tumor models. This technique requires the transmission of a train of pulses with varying amplitude, frequency, and phase within the train. High-amplitude and low-frequency pulses are applied to destroy the agent, and low-amplitude higher-frequency pulses are used to monitor the refresh of contrast agents into the sample volume. The signal processing techniques applied following data acquisition utilize the subharmonic component of the received signal. Mechanistic studies of contrast agent destruction, a Rayleigh–Plesset based signal model, and experimental data acquired from a rat tumor model will be included in our presentation. In addition, studies of the local delivery of drugs using oil-based drug delivery vehicles have been conducted. Results of these studies demonstrate that the probability of agent fragmentation and the size of the resulting fragment can be controlled using ultrasonic parameters. [We acknowledge the support of NIH CA 76062.]

2:45

2pBB3. Equilibrium loading of cells with molecules by ultrasound: Effects of molecular size, acoustic energy, and cell-to-bubble ratio. Mark Prausnitz, Hector Guzman, and Robyn Schlicher (School of Chemical Eng., Georgia Tech, Atlanta, GA 30332-0100)

Ultrasound has been shown to disrupt cell membranes and thereby load cells with molecules. In this study, we quantified levels of cellular uptake of calcein (0.6 kDa), bovine serum albumin (66 kDa), and two dextrans (42 and 464 kDa) into DU145 prostate cancer cells using 500 kHz ultrasound in the presence of Optison contrast agent. The results demonstrated that levels of uptake (i) were similar for all four molecules at the same ultrasound conditions and (ii) for some cells reached thermodynamic equilibrium with the extracellular solution. Over a broad range of acoustic conditions, these levels of uptake, as well as loss of cell viability, were shown to correlate with acoustic energy exposure. In a separate set of experiments, increasing cell concentration was shown to increase both viability and uptake, and increasing Optison bubble concentration decreased viability with no significant effect on uptake. As a parameter that can unify these two sets of data, the cell-to-bubble ratio was found to correlate well with both cell viability and the number of cells with uptake normalized relative to the number of Optison bubbles provided, which can be viewed as a measure of bubble ''efficiency.''

3:05–3:20 Break

3:20

2pBB4. Three variations on a theme: Ultrasound permeabilization of biologically interesting membranes. Pierre D. Mourad (Ctr. for Medical Acoust., Appl. Phys. Lab. and Dept. of Neurological Surgery, Univ. of Washington, Seattle, WA 98195)

This presentation focuses on three applications of therapeutic ultrasound for the purpose of permeabilizing membranes, with an emphasis, where possible, on the biology as well as the physics of the processes involved. In the first instance, ultrasound successfully and transiently opens the blood-brain barrier to the flux of blood-borne chemical dyes in vivo, as documented grossly, and with light and electron microscopy. In the second example, ultrasound transiently disrupts the membrane coating of a novel biopolymer tablet *in vitro* in order to release insulin carried within the tablet. The membrane consists of a series of self-assembling monolayers. In the final example, ultrasound acts synergistically with a novel polymer to permeabilize model lipid bilayers. The novelty here is that the effect is pH-dependent, due to the pH-dependence of the conformation of the polymer, and, that the polymer can be both targeted to cell surfaces with appropriate receptors, and can carry drugs or DNA. [Work supported by NIH and DARPA. I would like to thank my colleagues at CIMU and the Department of Neurological Surgery for their support in this work.]

3:40

2pBB5. Therapeutic applications of microbubbles. Evan C. Unger, Terry O. Matsunaga, Thomas P. McCreery, Robert Sweitzer, Patricia Schumann, Rachel Quigley, and Reena Zutshi (ImaRx Therapeutics, Inc., 1635 E. 18th St., Tucson, AZ 85719)

While microbubbles have been developed for diagnosis as ultrasound contrast agents, the greatest potential of these agents may lie in therapy. Microbubbles have extensive applications for treating vascular thrombosis, drug and gene delivery. Targeted microbubbles containing ligands for cell or site-specific recognition have been developed. These agents function like ''smart bubbles'' to home in on their targets. One such agent, MRX-408, a microbubble bearing targeting ligands directed to the activated receptor of platelets has undergone extensive pre-clinical testing. Recent studies comparing Sonothrombolysis with MRX-408 and nontargeted microbubbles show significantly more clot lysis with targeted as compared to nontargeted microbubbles. The targeted bubbles effectively concentrate cavitation nuclei within the clot. Additional targeted microbubbles for drug and gene delivery are also under development.

4:00

2pBB6. Sonodynamic treatment with cavitation-promoting agent activated by second-harmonic superimposed progressive waves. Shin-ichiro Umemura, Ken-ichi Kawabata, Nami Sugita, and Kazuaki Sasaki (Central Res. Lab., Hitachi, Ltd., Kokubunji, Tokyo 185-8601, Japan, sumemura@crl.hitachi.co.jp)

Sonodynamic therapy has been proposed based on the finding that certain chemicals such as hematoporphyrin are activated by acoustic cavitation and thereby induce a significant antitumor effect. It has also been found that sonodynamically active cavitation can be induced at a relatively low ultrasonic intensity through superimposing the second harmonic onto the fundamental. An *in vivo* as well as an *in vitro* cavitation threshold was an order of magnitude lowered by second-harmonic superimposition especially when combined with a cavitation-promoting sonochemically active agent such as rose bengal. Amphiphilic derivatives of rose bengal were synthesized so as to obtain tumor accumulating agents of this kind for sonodynamic tumor treatment. The tissue distribution of the derivatives in mice was investigated and those having a normal alkylic chain with 10–15 carbon atoms were proved to accumulate well in tumor tissues. With an administration of one of the derivatives, experimental murine tumors were treated with second-harmonic superimposed progressive waves at 0.5 and 1 MHz. The tumor growth was significantly suppressed by sonodynamic treatment in this combination while ultrasound alone showed only a slight effect. [Work supported in part by the National Research and Development Program for Medical and Welfare Apparatus, Japan.]

4:20–4:25
Break

4:25–5:25
Panel Discussion

5:25–5:30
Closing Remarks

TUESDAY AFTERNOON, 4 DECEMBER 2001 ROOM 220, 1:30 TO 3:45 P.M.

Session 2pEA

Engineering Acoustics: Sensors and Sources in Air

Ahmet Selamet, Cochair
Center for Automotive Research, The Ohio State University, 930 Kinnear Road, Columbus, Ohio 43212

Louis N. Cattafesta III, Cochair
Department of Aerospace Engineering, Mechanical and Engineering Sciences, University of Florida, Box 116250, Gainesville, Florida 32611

Contributed Papers

1:30

2pEA1. System for measurement of microphone distortion and linearity at very high sound levels. Erling Frederiksen (Brüel & Kjaer, 2850 Naerum, Denmark)

Even if sound measurements are made over a very wide dynamic range, say from 0 dB to 180 dB, the applied instruments are typically tested only at one level, which is most often within the range between 94 dB and 124 dB. Is this enough? Some types of microphone have probably never been tested at medium and high levels, as neither standardized nor generally agreed methods are described or available. However, today aerospace organizations, accreditation bodies and laboratories performing pattern approval of instruments ask for documentation, methods, and test equipment. Therefore, B&K decided to develop a system for generation of sound pressure over a wide dynamic range for routine testing of measurement microphones and pressure transducers. This paper describes the design and the properties of the system, which generates high pressures by a system of coupled tubes with standing waves. The general operation principle is not new, but the system, which can operate upon an ordinary laboratory table, is refined and optimized for a generation of levels as high as 174 dB at 500 Hz with a very low distortion (attention is paid to third harmonic distortion, which indicates lack of linearity).

1:45

2pEA2. Measurement of background noise in piezoelectric and electret condenser microphones. Allan J. Zuckerwar (NASA Langley Res. Ctr., M.S. 493, Hampton, VA 23681) and M. Roman Serbyn (Morgan State Univ., Baltimore, MD 21251)

An upgraded version of a previously described acoustic isolation vessel [K. C. T. Ngo and A. J. Zuckerwar, J. Acoust. Soc. Am. **93**, 2974–2980 (1993)] has been used to investigate the background noise in piezoresistive and electret condenser microphones over the frequency range 1 Hz–100 kHz. The test microphones represent four different manufacturers. Theoretical models for predicting the noise characteristics of piezoresistive (nonreciprocal) and electret condenser (reciprocal) microphones are proposed. In particular, the results of 1/f noise measurements on these microphones are compared with those previously obtained from air condenser microphones.

2:00

2pEA3. Technology development for directional acoustic arrays. David P. Arnold, Toshikazu Nishida (Dept. of Elec. and Computer Eng., Univ. of Florida, Box 116130, Gainesville, FL 32611), Louis N. Cattafesta III, and Mark Sheplak (Dept. of Aerosp. Eng., Mech., and Eng. Sci., Univ. of Florida, Box 116250, Gainesville, FL 32611)

The need for noise source localization and characterization has driven the development of advanced sound field measurement techniques using microphone arrays; however, the cost and complexity of these systems limit their widespread use. One common application is the study of aircraft noise in wind tunnels. The arrays typically consist of moderate numbers (<40) of conventional condenser microphones and range in size from several inches to several feet. Arrays consisting of more microphones and/or the use of multiple arrays may be needed to analyze complex multipole noise sources. However, the feasibility of scaling the current technology to multiple arrays with large numbers of microphones is limited by the cost per channel, array mobility, and data handling efficiency. A reduced cost, mobile, high speed 16-channel array has been constructed to demonstrate key technologies to address these scaling issues. The system is based on a hybrid package that combines a silicon piezoresistive microphone and amplification circuitry. Each package is socketed into a printed circuit board, providing electrical connections and structural support. High speed data acquisition and signal processing hardware is used to generate the array response in near real time. The results of the initial array calibrations are presented.

2:15

2pEA4. Effects of temperature and humidity on the bending stiffness of a sheet of paper measured by laser ultrasonics. Matthew A. Cornwell and Yves H. Berthelot (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

Laser ultrasonics has great potential for monitoring the bending stiffness of paper during on-line operations. The temperature and humidity conditions at the paper have a significant effect on the bending stiffness, and there is a need to quantify this dependence before an on-line system can be deployed. A laser pulsed source illuminates a point on a paper sample placed inside a controlled environmental chamber and generates ultrasound that is detected by a fiberoptic interferometer. The waveform is analyzed to extract the frequency dependence of the group velocity of the A0 lamb mode, and the corresponding bending stiffness. Three different methods of signal analysis are compared: the complex wavelet transform, the short-time Fourier transform, and the phase-unwrapping method, which requires measurements at two locations. The error bars associated with the bending stiffness are estimated both for the intrinsic experimental errors and for the point to point variablity in the paper itself. Results are presented for temperature ranging from 25 °C to 93 °C, and for relative humidity ranging from 2% to 75%. [Work supported by the Department of Energy and the Institute of Paper Science and Technology.]

2:30

2pEA5. A physically attractive alternative to Lighthill's quadrupole aerodynamic sound source. Alan Powell (Dept. of Mech. Eng., Univ. of Houston, TX 77204-4006)

Now an impressed force/unit vol. f_i acting *alone* on an acoustic medium results either directly in the formal mathematical monopole distribution $\partial f_i/\partial y_i$ but which is physically unrealistic, not appearing in the continuity equation, or the given dipole distribution f_i by the divergence theorem. These both yield point dipoles, $\mathcal{I}y_k\partial f_i/\partial y_i$ by Taylor expansion or the expected $\mathcal{I}f_i$, where $\mathcal{I}=$''*of strength proportional to the instantaneous volume integral of*.'' Aerodynamic sound generation based on Lighthill's inhomogeneous wave equation yieds directly a source distribution of the form (a) monopole $\partial^2 T_{ij}/\partial y_i\partial y_j$, $\approx\partial^2\rho u_i u_j/\partial y_i\partial y_j=\partial/\partial y_i \cdot \partial\rho u_i u_j/\partial y_j$. By the divergence theorem alternative *integrally equal* distributions are (b) dipole $\partial\rho u_i u_j/\partial y_j$ and (c) Lighthill's quadrupole $\rho u_i u_j$. All yield point quadrupoles for (a) $\mathcal{I}y_k y l\partial^2\rho u_i u_j/\partial y_i\partial y_j$ and (b) $\mathcal{I}y_k\partial\rho u_i u_j/\partial y_i$ by Taylor expansions, (c) $\mathcal{I}\rho u_i u_j$. Distribution (a) is unrealistic physically just like $\partial f_i/\partial y_i$; (b) appears *directly* in the momentum equation, like f_i, and is equivalent to an applied force/unit volume resulting in sound-generating fluid acceleration, while (c) appears in it only in derivative form. Hence (b) is an attractive alternative to the customary (c) because of that basic relationship which has simple physical appeal.

2:45–3:00 Break

3:00

2pEA6. Helmholtz resonator with an extended neck. Iljae Lee, Ahmet Selamet (The Ohio State Univ., Ctr. for Automotive Res., 930 Kinnear Rd., Columbus, OH 43212, selamet.1@osu.edu), and Kevin V. Tallio (Ford Motor Co.)

The present analytical, computational, and experimental work investigates the effects of length and shape of an extended neck on the acoustic attenuation performance of a concentric circular Helmholtz resonator with a fixed cavity. For an extension with constant cross-sectional area, a 2-D analytical approach is developed to determine the acoustic behavior, while a 3-D boundary element method is employed when the cross-sectional area is allowed to vary. These results are compared with: (1) the experimental data obtained from an impedance tube setup, and (2) the classical lumped analysis to illustrate the effects of wave propagation and higher-order modes. The impact of extension and its geometry on shifting the resonance frequency is emphasized.

3:15

2pEA7. The performance of two-dimensional acoustic resonators with large openings. K. Scott Line, Michael J. Anderson, and Ralph S. Budwig (Dept. of Mech. Eng., Univ. of Idaho, Moscow, ID 83844-0902)

We address the design of acoustic resonators with large openings. This problem is of interest to those who wish to achieve high levels of acoustic amplification in spite of energy leakage through the openings. A 2-D analytical model was used to predict the performance of a rectangular resonator with rigid walls and compare the performance to the corresponding situation for a cavity between two rigid semi-infinite walls. The fluid in the resonator was air, and the dissipation model included body absorption and simplified boundary layer losses. It was found that the collimation of the transducer field, and the nature of the attenuation model determined the performance of a resonator with large openings. For the dissipation model used, it was predicted that optimal performance of a resonator with openings would be obtained at a transducer ka of approximately 100. Experimental measurements of closed and open cavity quality factors, for transducer ka ranges lower than 100, showed the same trends as analytical predictions, however, the measured quality factors were much lower than the analytical predictions.

3:30

2pEA8. Inflatable lightweight low-frequency loudspeaker. Robert Corsaro (Code 7135, Naval Res. Lab., Washington, DC 20375)

In an aerospace application we required a low-mass sound generator with good performance at low audio frequencies (i.e., 30–300 Hz). One approach which has been successful in meeting these needs makes use of an inflatable tube. The actuation component is a ''Thunder'' actuator (Face Intl. Corp.), selected for its high displacement and fast response. Conventionally, when coupled to loudspeaker diaphragms these actuators have failed to produce the combination of force and displacement needed for low-frequency sound generation. In the present device the actuator is coupled instead to a flat, compliant diaphragm supported on the edges by an inflatable tube. Sound pressure levels measured in the farfield (25 cm) using only 200 V peak drive (one-third or its rating) we found nominal levels of 746 dB over the band from 38 to 330 Hz. In the uninteresting region above 330 Hz the output is somewhat higher, but remains well behaved to at least 2 kHz. [Work supported by the NRL/ONR Smart Blanket program, which is exploring the use of active controlled surface covers for reducing sound levels in payload fairing regions.]

Session 2pMU

Musical Acoustics: Caribbean Musical Instruments and Traditions

Paul A. Wheeler, Cochair
1595 North 1600 East, Logan, Utah 84341

Thomas D. Rossing, Cochair
Physics Department, Northern Illinois University, DeKalb, Illinois 60115

Chair's Introduction—1:00

Invited Papers

1:05

2pMU1. Overview of Caribbean music. Martha Ellen Davis (School of Music, Univ. of Florida, 1350 Turlington, Gainesville, FL 32611)

To frame the 2001 meeting of the ASA within the soundscape of the region, this paper offers an overview of the diversity of musical sounds and sound-makers, i.e., instruments, of the Caribbean and Circum-Caribbean region. Many are in fact present in the Miami–Dade/Fort Lauderdale–Broward area because of the growing presence of immigrant populations, as well as the influence of some of their traditions on ''world beat'' music (the subset of ''world music'' with African influence). The survey of sounds of the Caribbean, illustrated audiovisually, will include both the commercially disseminated genres (''popular music''), and others which are not marketed or marketable, although in some cases the roots of popular music. The ''unknown'' genres include social dances, as well as work songs, sacred drumming, ritual songs, and more. The paper proposes African culture influence as the common denominator which musically unifies the region and explains why.

1:35

2pMU2. Traditional Junkanoo music instruments of the Bahamas. Nina Wood (P.O. Box 15864, Plantation, FL 33318-5864)

Junkanoo is a carnival-like parade that takes place twice a year in the Bahamas. The parade grew out of the music practices of Africans who had been enslaved in the Bahamas during the nineteenth century. Today, the parade incorporates large, elaborate costumes, and music played on traditional instruments. The number of participants in Junkanoo has grown tremendously, and moves have been made to incorporate amplified, electronic music in the parade in order to project the volume of the music over the sounds of the participants and spectators. Despite attempts to incorporate amplified music, however, acoustic instruments have remained primary. Junkanoo music performance has retained many of its African characteristics in terms of timbre, percussion, and density, and throughout the past two decades adjustments have been made to the instruments to enhance these characteristics. In this discussion the instruments of the Junkanoo music tradition will be introduced, their timbral qualities will be examined, and their particular functions and relationships within the tradition will be explored.

2:05

2pMU3. Science of steelpans: What is known and what is not. Thomas D. Rossing (Phys. Dept., Northern Illinois Univ., DeKalb, IL 60115) and Uwe J. Hansen (Indiana State Univ., Terre Haute, IN 47809)

One of the most important musical instruments developed in the 20th century is the Caribbean steelpan. In addition to being the foremost musical instrument in its home country of Trinidad and Tobago, the steelpan is becoming increasingly popular in North America and Europe. Scientific studies on the instrument have not kept pace with its growing popularity, however. This paper reviews some of the scientific studies that have been made on steelpans in our laboratory and elsewhere. One of the most important steps in understanding the steelpan is to understand how it vibrates and produces sound. We report measurements of modal shapes in different steelpans by prominent makers and attempt to interpret these results in terms of the geometry and fabrication history of these pans. We also discuss the physics of the material, including the residual stresses, the surface hardness, and the effect of heating.

2:35

2pMU4. Modal analysis of steelpans: Holographic, computer, sand patterns, and microphone scannning. Uwe J. Hansen (Dept. of Phys., Indiana State Univ., Terre Haute, IN 47809) and Thomas D. Rossing (Dept. of Phys., Northern Illinois Univ., DeKalb, IL 60115)

We compare several methods for modal analysis that have been applied to Caribbean steelpans. Holographic interferometry offers the best spatial resolution of operating deflection shapes, and hence of normal modes. Recording holographs on photographic film tends to be rather time consuming, however. Electronic TV holography, on the other hand, offers the opportunity to observe vibration

motion nearly in real time. Striking a steelpan with an instrumented force hammer excites many modes at once, and a computer program is used to determine the normal modes by means of a curve-fitting program. In the oldest method of modal analysis (and still a useful one), a fine powder is sprinkled on the surface of a steelpan vibrating at the frequency of a given mode. The powder collects at the nodal lines, providing a Chladni pattern or map of that mode.

3:05–3:15 Break

3:15

2pMU5. The Steelpan Development Centre at The University of the West Indies. Brian Copeland, Clément Imbert, and Derek Gay (Faculty of Eng., The Univ. of the West Indies, St. Augustine, Trinidad and Tobago, West Indies)

This paper describes the Steelpan Development Centre (SPDC) of The University of the West Indies (UWI), a specialist center created to explore, support, and develop all aspects of the technology of the steelpan. The SPDC places special emphasis on the development on steelpan manufacturing technology and performance in the Caribbean region. While the major focus of the center is pan technology, its mission also includes consideration of other aspects that impact on the growth of the instrument, in collaboration with the UWI Center for Creative and Festival Arts. In particular, the SPDC has strong interest in providing technical, business, and administrative support to the many regional tuners who, by and large, still operate at the craft level. The paper looks at the motivation for the formation of the center, a few of the developments to date and some of the projects listed for consideration.

3:45

2pMU6. Materials and manufacturing of steelpans. C. A. C. Imbert (Steelpan Development Ctr., Faculty of Eng., The Univ. of the West Indies, St. Augustine, Trinidad and Tobago, West Indies)

The musical instrument referred to as the Caribbean steel drum (called the steelpan or simply pan by true aficionados) is traditionally hand-crafted from 55 U.S. gallon oil drums. These drums are made to satisfy certain physical and mechanical properties but certainly not for musical instruments. As a result pan builders/tuners have to reject a relatively large number of drums because they do not respond well enough to produce fine quality lead instruments. In this paper the results of the mechanical/metallurgical investigation of several drum steels are reported—steels that respond well and others that do not—as well as the process of hardening steelpans by nitriding. Also reported are the results of mechanized methods of sinking the pan compared to the traditional manual process.

Contributed Paper

4:15

2pMU7. Sound radiation from Caribbean steelpans. Brian Copeland (Dept. of Elec. and Computer Eng., Univ. of the West Indies, St. Augustine, Trinidad), Andrew Morrison, and Thomas D. Rossing (Phys. Dept., Northern Illinois Univ., DeKalb, IL 60115)

Besides radiating sound from the note area being struck, a steelpan radiates from neighboring note areas that vibrate sympathetically, from the areas between notes, and from the skirt [Rossing *et al.*, Phys. Today **49**, 24–29 (1996)]. We have used a two-microphone intensity probe to measure the sound field around steelpans when various note areas are excited with swept sine wave force in an anechoic room. The radiation patterns in planes parallel and perpendicular to the surface of the pan are plotted and analyzed.

Session 2pNS

Noise: Noise Measurement, Sources, Control

Joseph Pope, Cochair
Pope Engineering Company, P.O. Box 236, Newton, Massachusetts 02459-0002

Bennett M. Brooks, Cochair
Brooks Acoustics Corporation, 27 Hartford Turnpike, Vernon, Connecticut 06066

Contributed Papers

1:30

2pNS1. Directivity of the diffracted field around a half-plane. Jeong Ho You, Penelope Menounou, and Wayne M. Wright (Dept. of Mech. Eng., Univ. of Texas, Austin, TX 78712)

The directivity of the diffracted field around a half-plane barrier is investigated. Experiments were carried out in air using a spark source generating N-shaped pulses, a capacitor microphone, and a rigid aluminum plate with: (i) straight top edge (reference case), (ii) jagged top edge, and (iii) straight edge covered with sound absorptive material. Measurements were taken around the plate by varying the angular position of the microphone while keeping its radial distance from the barrier edge constant. The following were observed: (i) Straight edge: The diffracted pulse changes polarity as it passes across the two shadow boundaries and the continuation of the barrier in the half-space above it. (ii) Jagged edge: Although jaggedness in the top edge of the barrier drastically changes the shape of the diffracted pulse, the directivity pattern remains roughly the same. (iii) Sound absorptive edge: Although the shape of the diffracted pulse remains the same as in the reference case, the directivity patterns change considerably and changes are more pronounced at the source side. The experimental data and observations were in good agreement with theoretical predictions from the Directive Line Source Model [Menounou *et al.*, J. Acoust. Soc. Am. **107**, 103–111 (2000)]. [Work supported by Texas ATP.]

1:45

2pNS2. Feasibility of a phased acoustic array for monitoring acoustic signatures from meshing gear teeth. Adrian A. Hood and Darryll J. Pines (Univ. of Maryland, College Park, MD 20143, djpterp@eng.umd.edu)

This research investigates the feasibility of sensing damage emanating from rotating drive train elements such as bearings, gear teeth, and drive shafts via airborne paths. The use of a planar phased acoustic array is evaluated as a potential fault detection scheme for detecting acoustic signatures radiating from gearbox components. Specifically, the use of beam focusing and steering to monitor individual tooth mesh dynamics are analyzed, taking into consideration the constraints of the array/gearbox geometry and the spectral content of typical gear noise. Experimental results for a linear array are presented to illustrate the concepts of adaptive beam steering and spatial acoustic filtering. Also, a comparison is made between the radiated spectrum of a damage free system and one with seeded faults using spur gears. This feasibility study indicates that the planar array can be used to track the acoustic signatures of gear mesh dynamics of typical helicopter systems at higher harmonics of the mesh frequency.

2:00

2pNS3. Machine vision improves sound level meter calibration. Guillermo Arcas, Mariano Ruiz, and Raul Pagan (Universidad Politecnica de Madrid, INSIA Crta. Valencia Km-7, Madrid 28031, Spain)

In this paper, important improvements in sound level meter (SLM) calibration are presented. One of the most important problems that the calibration laboratories are facing today is the long time required for calibrating certain SLMs. Instruments without a dc output require that an operator reads the SLM response in every test. This makes the process slow and tedious, and induces errors due to the operators fatigue. By using an image acquisition card with a low cost analog camera to acquire an image of the sound level meters display, calibration time can be reduced significantly. Once an image of the SLMs display has been acquired, machine vision techniques (OCR) are used to recognize the sound level meter indication. This process reduces calibration time and errors produced by operators fatigue.

2:15

2pNS4. The software architecture for an analyzer sound level meter with DSP-bios. Juan M. Lopez Navarro, Mariano Ruiz Gonzalez, Manuel Recuero Lopez, and Raul Pagan Muoz (INSIA Ctra. de Valencia Km 7, Madrid 28031, Spain)

This paper shows the software architecture implemented for the accomplishment of an analyzer sound level meter in one-third octave based on a Digital Signal Processor TMS320C6701 of Texas Instruments. In order to minimize the cost of the equipment, it has been done without the elements of interface with the user (keyboard and displays). For the operation and control of the equipment a laptop, or a PDA may be used. A communication RS232C is used in the local way; and in a remote way, a connection through built-in MODEM GSM. The sever application is executed in the signal processor. It has been implemented in C language and it runs on an real time operating system (DSP-Bios provided by Texas Instruments). DSP-Bios provides the management of events, queues, and timings. The application has the necessary algorithms to make the measures and sends the results to the visualizer, or to a center of remote operation by means of the GSM. The client application that is executed in the visualization element (Laptop or PDA) is developed in Visual Basic. This application can run on any Windows operating system depending on the terminal used: Windows 9x or Windows 2000 for portable ones and on Windows CE for mobile ones.

2:30

2pNS5. Determination of reference sound source sound power levels using ARI Standard 250. Stephen J. Lind (The Trane Co., 3600 Pammel Creek Rd., La Crosse, WI 54601)

The Air Conditioning and Refrigeration Institute (ARI) Standard 250 Performance and Calibration of Reference Sound Sources was used to establish one-third octave band sound power levels of several reference sound sources (RSS). The work was carried out in an anechoic chamber that was qualified at frequencies of 100 Hz and above using ANSI S12.35. Measurements were made with circular slices around the RSS at heights from 5 cm to 195 cm in 10 cm increments using both intensity and sound pressure. A 19 point fixed array was also employed for comparison of one of the sources. Sound power levels were compared from the 50 Hz to the 10 000 Hz one-third octave bands using several methods. The results are compared to the sound power levels provided by the manufacturer.

2:45–3:00 Break

3:00

2pNS6. Numerical prediction of noise generated from low-pressure axial-flow fans. Li Song and Wang Dongbao (Dept. of Eng. Mech., Tsinghua Univ., Haidian District, Beijing 100084, PROC)

Low-pressure axial-flow fans are widely used in ventilation, air conditioning, and cooling, but it is very difficult to predict their noise. In 1993 Lee *et al.* predicted the noise level of low-pressure axial fans according to a wake-flow shedding model based on the mechanism that the noise is induced by a lift force pulse caused by wake-flow vortex shedding [J. Sound Vib. **164**, 327–336 (1993)]. The main approximation that results in considerable errors is that the parameters of a flat-plate boundary layer at the trailing edge are adopted instead of that of a blade boundary layer. In this paper the parameters of the boundary layer, including the distribution of nominal thickness and momentum thickness of boundary layer along the blade height, are obtained by numerical simulation of the 3-D turbulent flow field in axial-flow fans. It is shown from the predicted results for two model axial fans that the noise level agrees with measured data quite well. The errors are 2–4 dB for A-weighted sound pressure level (SPL) and 3–4 dB for linear SPL. Furthermore the effects of parameters of the blade boundary layer on noise level and spectrum are also discussed in more detail in this paper.

3:15

2pNS7. Source models for turbulence generated sound in ducts. Mats Abom (MWL, Dept. of Vehicle Eng., KTH, 100 44 Stockholm, Sweden)

Turbulence or flow generated sound is a problem in many duct and pipe systems. This type of noise will add to the primary noise from fluid machines (e.g., fans) connected to the system and often sets the limit for the sound attenuation possible. For low Mach numbers the flow generated noise will primarily be associated with unsteady flow separation occurring, e.g., at constrictions and bends, resulting in a dipole type of source. In this paper a procedure for deriving source models for turbulence generated sound in the low-frequency plane wave range is described. The source models are formulated as active acoustic 1- and 2-ports [H. Boden and M. Abom, Acta Acustica **3**, 549–560 (1995)], which makes them suitable for implementation in software for analysis of duct networks. Regarding the active part or the source strength, it can be obtained using a semi-empirical model for flow separation noise [P. A. Nelson and C. L. Morfey, J. Sound. Vib. **79**, 263–289 (1981)]. The passive part can be well approximated in the low-frequency limit by a quasi-stationary model derived from the conservation of mass and momentum in the flow. Examples of tests and validation of the proposed model are also presented.

3:30

2pNS8. Broadband passive noise control in air/water using structural acoustic silencers. Sripriya Ramamoorthy, William P. Cook, Karl Grosh (Dept. of Mech. Eng., Univ. of Michigan, 2350 Hayward Ave., Ann Arbor, MI 48109-2125), and Tony G. Nawar (Dept. of Elec. Eng. and Computer Sci., Univ. of Michigan, 1301 Beal Ave., Ann Arbor, MI 48109-2122)

The effectiveness of introducing flexible structural layers into fluid and air conveying ducts for controlling noise is investigated through theoretical and experimental means. Previous work has shown that using flexible rather than rigid walls can theoretically achieve high transmission losses. However, for hydraulic applications these relatively thin layers cannot sustain the high static pressures seen in real situations and for pneumatic applications, radiated noise from these walls (called breakout noise) is a significant issue. Designs that introduce a second duct below the flexible wall region of the main duct allows for equilibration of the static pressure and elimination of breakout noise. However, plane waves can now propagate through the duct without exciting the plate, thus reducing the effectiveness of the device. Designs suitable for hydraulic/pneumatic applications, that overcome these issues and achieve appreciable transmission loss are investigated. Results based on 2.5 D finite element simulations will be compared with experimental results. [Research funded by ONR and NSF.]

3:45

2pNS9. The effect of active noise control actuator placement on the total sound power radiated by modally vibrating cylinders and plates. Jennifer L. Cooper and Jiri Tichy (Grad. Prog. in Acoust., The Penn State Univ., University Park, PA 16801)

Many practical radiators can be modeled as vibrating bars or plates. The radiation caused by their modal vibration is studied replacing them by point sources at each antinodal location. Active control of their radiated sound power is investigated using point secondary sources. The maximum obtainable attenuation is examined as a function of the distance from primary to secondary sources and the separation of secondary sources. For a given distance from the primary sources, the obtained attenuation is independent of secondary source separation for a range of wavelengths. For larger secondary source separations, the control can be simplified, allowing each secondary source to be controlled by only one microphone, as opposed to the usual method of controlling all secondary sources by all the microphones.

4:00

2pNS10. Airfoil vibration in the wake of an obstacle: Effects of inflow nonuniformity and vortex sheet. Svetlana Kovinskaya, Eduard Amromin (Mechmath LLC, 2109 Edmond, OK 73034), and Genrietta Avilova (Andreev's Acoust. Inst., Moscow, Russia)

Obstacles in flow produce wakes that can cause significant vibration of devices located within the wakes. Such vibration appears in many situations, from landing an aircraft on a ship deck to operations of rotating machinery. There are two principal singularities of these wakes that define incoming flow for the downstream devices. First, the time-average velocity profiles are highly nonuniform. Second, flow separation from obstacles is accompanied by generation of large vortices. An initial mathematical study of this flow-induced vibration of the devices includes quasipotential approach in determination of time-averaged flow parameters and determination of amplitude response of elastic airfoils (device cross sections) on perturbation. Perturbation frequencies depend on the obstacle drag, size and free-stream speed. Vortex sizes depend also on Reynolds number. Numerical modeling shows a certain difference between the foil response on gust flow and its response on excitations caused by the obstacle wakes.

Session 2pPAa

Physical Acoustics: High Amplitude Effects in Resonators II

Anthony A. Atchley, Cochair
Graduate Program in Acoustics, Pennsylvania State University, P.O. Box 30, Applied Science Building, State College, Pennsylvania 16804

Victor W. Sparrow, Cochair
Graduate Program in Acoustics, Pennsylvania State University, P.O. Box 30, Applied Science Building, State College, Pennsylvania 16804

Invited Papers

1:00

2pPAa1. Acoustic agglomeration experiments and modeling. G. Douglas Meegan, Chris E. Peterson, and Ronald E. Kumon (Appl. Res. Labs, Univ. of Texas, Austin, TX 78713-8029)

Acoustic agglomeration of aerosols has been known since at least 1931 when it was observed that particles tend to ''stick'' together in the presence of an intense acoustic field, thereby forming larger particles. Since the 1970s acoustic agglomeration has been an active area of research because of its potential applications for the control of particulate emissions resulting from fossil fuel combustion. In this talk, experimental methods used in the past to study acoustic agglomeration are described, along with improvements we have introduced. A simple standing wave tube is one of the devices used in our laboratory to measure the dependence of acoustic agglomeration on sound pressure level, frequency, and exposure time. Results are presented and compared with theoretical predictions based on existing models. The potential application to diesel engine emissions and the filtration of hazardous particulate will be discussed.

1:30

2pPAa2. A computational study of acoustic streaming in a duct. William R. Erskine and Ahmet Selamet (Dept. of Mech. Eng., The Ohio State Univ., Columbus, OH 43210)

The present work is a continuing investigation of Rayleigh streaming, which is a nonlinear phenomenon characterized by acoustically driven flows attenuated by viscous effects due to the presence of walls, resulting in nonzero average streaming velocities and circulation zones within a duct. A 2-D long and narrow duct is considered with one end closed and the other end subject to a sinusoidal acoustic velocity at quarter-wavelength frequencies and varying amplitudes, leading to a standing wave within the duct. A computational fluid dynamics code based on the Pressure Implicit with Split Operators (PISO) algorithm and finite volume method is used to solve the unsteady, compressible Navier–Stokes equations with turbulence modeling. Since the formation of streaming depends on the velocity profile within the Stokes layer, the grid resolution is increased near the walls. The effect of inlet conditions is studied in terms of circulation patterns, including the number of zones per wavelength, penetration depth into the fluid, and mean velocity profiles across the boundary layer. Some of the computational predictions are also compared with the classical analytical treatments.

Contributed Papers

2:00

2pPAa3. Shock-free, forced oscillations of an air column in a closed tube. Nobumasa Sugimoto, Takeshi Hashiguchi (Dept. of Mech. Sci., Grad. School of Eng. Sci., Univ. of Osaka, Toyonaka, Osaka 560-8531, Japan), and Mitsuhiro Masuda (Dept. of Thermal Energy Systems, Ecology and Energy Systems Res. Ctr., Sanyo Electric Co., Hashiridani, Hirakata, Osaka 573-8534, Japan)

This paper reports generation of high-amplitude, shock-free oscillations of an air column in a closed tube of uniform cross section. The air column is driven sinusoidally by the bellows mounted at one end of the tube at a frequency equal or close to the resonance frequency of the lowest mode. The shock-free oscillations are achieved by connecting an array of Helmholtz resonators to the tube axially. The array of resonators gives rise to wave dispersion, which changes the sound speed in this tube to depend on the frequency and makes the tube *dissonant* by avoiding coincidence of the higher-harmonic frequencies of excitation and the resonance ones of the higher modes. Each resonator is identical and small in a sense that the cavity's volume is chosen much smaller than the tube's volume per axial spacing between neighboring resonators, while the natural frequency of the resonator is chosen even higher than the lowest resonance frequency. Frequency response is obtained and compared with the theory up to the resonance frequency of the third mode. It is observed that the shock-free oscillations are realized at the sound pressure level of 170 dB where the shock would emerge unless the array were connected.

2:15

2pPAa4. Suppression of nonlinear energy decrease by linear absorption and enhancement of resonant properties of nonlinear systems. Oleg Rudenko (Dept. of Acoust., Faculty of Phys., Moscow State Univ., Moscow 119899, Russia, rudenko@acs366b.phys.msu.su)

A general question of nonlinear vibration and wave theory is discussed which is of special interest for nonlinear acoustics and technologies. Examples of nonlinear systems of different physical nature are given to illustrate how the properly introduced additional absorption can decrease the nonlinear losses and improve the resonant properties of nonlinear systems. Both traveling and localized standing waves are considered. The possibilities to enhance the Q-factor of acoustic resonators are discussed in detail, as well as links to the turbulence, solid state physics, and electronics. [Work supported by CRDF, INTAS, and RFBR grants.]

Session 2pPAb

Physical Acoustics: Thermoacoustics

Richard Raspet, Chair
National Center for Physical Acoustics, University of Mississippi, 1 Coliseum Drive, University, Mississippi 38677

Contributed Papers

2:45

2pPAb1. Performance of a thermoacoustic prime mover without a stack. Stefan J. Turneaure, Ray Scott Wakeland, and Robert M. Keolian (Grad. Prog. in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804)

A standing-wave thermoacoustic prime mover which does not contain a stack has been constructed and tested. The device consists of two heat exchangers and a resonator body. The cold heat exchanger may be moved relative to the hot heat exchanger while the resonator is sealed and thermoacoustically generated sound is present. The device has a cross-sectional area of about 500 cm^2 and operates near 600 Hz with ambient air as the working gas. It will be explained how a no-stack thermoacoustic device may in principle have a greater efficiency than a stack-based device. In contrast to the well understood stack-based standing wave thermoacoustic engines, it is expected that the efficiency for a no-stack device will be nearly proportional to the acoustic pressure amplitude. Preliminary measurements on the device have achieved acoustic pressure amplitudes of 2.5% of the mean pressure. An interesting result is the presence of spontaneous thermoacoustic oscillations even when the gap between the cold and hot heat exchangers is several times larger than the gas displacement, $2x_1$. [Work supported by ONR.]

3:00

2pPAb2. Investigation of stack materials for miniature thermoacoustic engines. Ehab Abdel-Rahman, Young Kwon, and Orest G. Symko (Dept. of Phys., Univ. of Utah, Salt Lake City, UT 84112-0830, ehab@physics.utah.edu)

In a thermoacoustic engine one of the most critical components is the stack that interacts thermally with a sound field. Its effective surface area determines directly the cooling power and efficiency while the thermal properties determine the losses. Moreover, as the engines are miniaturized, stack performance on a small scale becomes a limiting factor, especially when a stack length of 1000 μ or less has to maintain temperature differences of around 30°C. We have investigated as stack material cotton wool, glass wool, ceramics, silica aerogel, carbon fibers, and carbon powder by measuring flow resistance at high acoustic frequencies, 4–20 kHz, and thermal transport properties. The mechanical properties and ease of machining to small dimensions, of the order of 1 μ, have been studied. In selecting the ideal material for very small devices, a compromise needs to be made between thermoacoustic and mechanical properties. [Work supported by ONR.]

3:15

2pPAb3. Experimental study of a thermoacoustic electrical generator. Jin Liu, Philip S. Spoor, John A. Corey (CFIC, Inc., Troy, NY 12180), and Steven L. Garrett (Penn State Univ., State College, PA 16804)

The stack and heat exchangers from the historic Five Inch Engine [G. W. Swift, J. Acoust. Soc. Am. **92**, 1551 (1992)] have been joined to a 2.02-m long resonator tube terminated by a variable acoustic load [A. M. Fusco *et al.*, J. Acoust. Soc. Am. **91**, 2229 (1992)]. Operation in 20-bar pure helium delivered 160 W to the load with an electrical heat input of 2.76 kW, producing a hot-end external duct temperature of 655 °C. Applied heat measuring 1.41 kW was exhausted through the cold exchanger to a water stream at 25 °C, with the remaining power producing useful work or otherwise lost by conduction through the engine and to the surroundings. Measured acoustic pressure amplitudes for electrical heat inputs in the range from 348 to 2773 W agreed with the DELTAE model to within 12%, but the error in the hot-end temperature prediction was 25%. The engine was then attached to a CFIC Model B-300 electrodynamic motor/alternator to explore feasibility of distributed power generation and co-generation. [The loan of the 5-in. engine, by G. W. Swift, Los Alamos National Laboratory, is gratefully acknowledged. J.L. and S.L.G. were supported, in part, by ONR.]

3:30

2pPAb4. Wet-walled thermoacoustics. William Slaton and Richard Raspet (The Natl. Ctr. for Physical Acoust., The Univ. of Mississippi, University, MS 38677)

An analytic solution of sound propagation in wet-walled tubes with a temperature gradient will be presented. The tube contains an inert gas–vapor mixture with a thin layer of condensed vapor coating the tube wall. The vapor phase condenses and evaporates from this layer during an acoustic cycle. This phased evaporation and condensation modifies traditional energy density and wave number equations. It is found that mass and heat transport act in parallel thereby increasing the acoustic energy density for proper choice of gas mixtures. This increase in energy density due to inert gas–vapor working fluids can be quite large and even dominate the traditional thermal transport contribution. The use of phased evaporation-condensation in gas mixtures provides a path to higher energy density in thermoacoustic systems. The effects on sound propagation, energy density, and efficiency due to the presence of an inert gas–vapor working fluid in tubes with wet walls will be discussed. Also, computational modeling of thermoacoustic refrigerators utilizing an inert gas–vapor working fluid will be presented. Modeling is done using DeltaE software with modifications made to explore inert gas–vapor mixtures. Design of a practical thermoacoustic refrigerator utilizing wet-walled thermoacoustics will be presented. [Work supported by ONR.]

3:45

2pPAb5. Reduced onset temperature difference in wet thermoacoustic engines. William V. Slaton, John W. Rayburn, Robert A. Hiller, and Richard Raspet (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

Measurements of the onset temperatures of a standing-wave thermoacoustic prime mover in air show a dramatically decreased onset temperature difference when liquid water is present on the stack surfaces. These data may be explained by considering the condensation and evaporation of

water during the acoustic cycle [W. V. Slaton and R. Raspet, J. Acoust. Soc. Am. **108**, 2569 (2000)]. The inert gas–vapor mixture thermoacoustic propagation equations are integrated through the resonator/stack system to determine the quality factor and resonant frequency using complex frequency analysis. In addition the 1-D thermoacoustic modeling code DELTA E was modified to include these effects in the stack to predict onset temperatures.

4:00

2pPAb6. The use of parallel-plate regenerators in thermoacoustic-Stirling engines. Scott Backhaus and Gregory W. Swift (Condensed Matter and Thermal Phys. Group, Los Alamos Natl. Lab., Los Alamos, NM 87545)

The regenerator is a key element in a thermoacoustic-Stirling engine. Typically, it is made by stacking up many layers of fine-mesh stainless-steel screen. With a high heat capacity and a typical pore size much smaller than the thermal penetration depth, it provides a nearly isothermal environment at the acoustic frequency. By almost eliminating the temperature oscillations of the gas in it, the regenerator minimizes the unwanted enthalpy flux from the hot heat exchanger to ambient heat exchanger and keeps the engine's efficiency high. However, the small pore size also causes a significant amount of viscous dissipation, which lowers the efficiency. A regenerator made of parallel plates can be designed to provide the same heat transfer properties as a screen regenerator with much lower viscous dissipation. By replacing a screen based regenerator with a carefully constructed parallel-plate regenerator, the acoustic power output of a thermoacoustic-Stirling engine is nearly doubled, with significant increases in efficiency at high amplitudes. [Work funded by the Office of Basic Energy Sciences within the U.S. DOE.]

4:15

2pPAb7. One-dimensional acoustic streaming instabilities in traveling-wave thermoacoustic devices employing jet pumps. Scott Backhaus and Gregory W. Swift (Condensed Matter and Thermal Phys. Group, Los Alamos Natl. Lab., Los Alamos, NM 87545)

Traveling-wave thermoacoustic devices employ a Stirling-type thermodynamic cycle and thus enjoy an inherent efficiency advantage over standing-wave devices. However, if the regenerator of the traveling-wave device is embedded in a loop, acoustic streaming around the loop can cause a heat leak that greatly reduces the efficiency. In several traveling-wave devices, asymmetric nozzles, or synthetic-jet pumps, have been successfully used to suppress the streaming and, therefore, the heat leak. However, in one traveling-wave refrigerator, an instability that precludes streaming suppression is driven by the interaction between the synthetic-jet pump and the thermoacoustic system. Calculations showing how this instability arises will be presented. [Work funded by the Office of Basic Energy Sciences within the U.S. DOE.]

4:30

2pPAb8. A study of thermoacoustic instabilities in a resonator with heat release. Konstantin I. Matveev and Fred E. C. Culick (Caltech, 301-46, Pasadena, CA 91125)

Excitation of acoustic modes in systems with heat release is a complex process, involving coupling between the heat transfer rate and the acoustic pressure. The Rijke tube is the simplest device for studying the fundamental principles of thermoacoustic instabilities both experimentally and theoretically. In our investigation, special attention is paid to reproducibility and precision of the measurements. Experimental results for transition to instability are obtained when controlled parameters (flow rate, heater position, and supplied power) are varied in a quasisteady fashion. The stability boundaries exhibit hysteretic behavior not previously reported. Properties of the limit cycles in the region of instability have also been determined. A mathematical model is developed that incorporates heat transfer, acoustics, and thermoacoustic interactions. An energy balance, accounting for all types of heat transfer, allows for the determination of temperature variations and heat flux along the tube. The acoustic modes, generally nonorthogonal, corresponding to particular experimental thermal conditions, are computed. Solutions of the linearized wave equation with both loss and heat source terms determine the stability of acoustic modes. Hysteresis effects at the stability boundary and amplitudes of the limit-cycles are calculated from the nonlinear model.

4:45

2pPAb9. Saturation of the thermoacoustic separation of a binary gas mixture. Drew A. Geller and Gregory W. Swift (Condensed Matter and Thermal Phys. Group, Los Alamos Natl. Lab., Los Alamos, NM 87545)

Spoor and Swift have previously shown [Phys. Rev. Lett. **85**, 1645–1649 (2000)] that sound waves can produce the separation of a binary mixture of gases in a duct. This phenomenon is caused by the conspiracy of two effects at the thermoviscous boundary layer: (i) oscillating temperature gradients drive thermal diffusion perpendicular to the direction of sound propagation; and (ii) viscosity sets up a time-averaged counterflow process. In order to assess the relevance and potential uses of this mechanism, it is important to determine the concentration gradient at which the separation process saturates. Here, it is shown that the saturation is determined primarily by ordinary diffusion and by a remixing flux due to the oscillatory motion. Both of these remixing processes are proportional to the concentration gradient, but only the latter process depends on the acoustics, being proportional to the squared volume flow rate and inversely proportional to the frequency. For He–Ar mixtures, saturation concentration gradients near 10%/m have been achieved, in agreement with detailed calculations of these processes. Evidence to date supports the possibility of more challenging separations, e.g., of isotopes. [Work supported by the DOE Office of Basic Energy Sciences.]

Session 2pPP

Psychological and Physiological Acoustics: Localization, Pitch, Hearing Aids, Instrumentation and Other Topics (Poster Session)

Constantine Trahiotis, Chair
Surgery and Neuroscience, University of Connecticut Health Center, L4071, Farmington, Connecticut 06032

Contributed Papers

All posters will be on display from 1:30 p.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:15 p.m. and contributors of even-numbered papers will be at their posters from 3:15 p.m. to 5:00 p.m.

2pPP1. The effects of presentation time on the localization accuracy of 3D sound through headphones. Fang Chen and Jonas Sääv (Linköping Univ. IKP/IAV, SE-581 83 Linköping, Sweden, fanch@ikp.liu.se)

The present study deals with the relationship between sound presentation time and localization accuracy of 3D sound presented through headphones. Huron Lake CP4 system was used for sound presentation and Flock of birds tracking system was used to register movement of the head. There were 31 volunteers (15 females and 16 males) in this experiment. Six different types of sounds were presented with the duration of 0.5, 2, 4, and 6 s. They were dog bark (0–500 Hz), coin drop with three different bandwidths (0–1650 Hz, 1650–6500 Hz, and 0–8500 Hz), alarm (0–5500 Hz) and male speech voice (0–7000 Hz). The results showed that the accuracy of localizing the sound is significantly related to the duration of the sound presentation. The perceived azimuth difference for 0.5 s duration ($45° \pm 15°$ in average) is significantly higher compared with the longer presentation duration. The perceived azimuth difference for 2 s duration is also significantly higher than 4 and 6 s duration ($21° \pm 9°$ in average). There is not any significant difference between 4 and 6 s duration. There is not any significant difference between male and female subjects, but individual difference is significant.

2pPP2. The effect of pinna protrusion angle on localization of virtual sound in the horizontal plane. Navarun Gupta, Carlos Ordonez, and Armando Barreto (Florida Intl. Univ., Dept. of Elec. and Computer Eng., 10555 W. Flagler St., Miami, FL 33174)

The transformation of a sound from its origin to each of the eardrums of a listener, due to the head, torso, and outer ear of the listener, are modeled by the so-called head-related transfer functions (HRTFs) for each individual and each location around the listener. These HRTFs can be used to transform a monoaural sound into left and right sounds that will give the subject the illusion of a virtual sound placement, at the location associated with the HRTF used. However, subjects tend to confuse sounds virtually placed in the front hemisphere with sounds placed symmetrically in the back hemisphere, i.e., points that lay in a cone of confusion. This paper reports on a study that involved the measurement of the HRTFs of 20 subjects, and the analysis of their performance in locating virtual sources in a horizontal plane. The subjects were tested using their own individual HTRFs and also using the HRTFs of a prototype subject who has particularly protruding pinnae. Results indicate that the additional shadowing of sounds from the back hemisphere encoded in the HRTFs of the prototype subject aided some of the test subjects in resolving the front–back confusion observed when tested with their own HRTFs.

2pPP3. Determining the influence of visual stimuli on the peception of surround sound using data mining algorithms. Piotr Odya, Andrzej Czyzewski, Bozena Kostek (Sound and Vision Eng. Dept., Technol. Univ. of Gdansk, 80-952 Gdansk, Poland), and Tomasz Smolinski (Computional Intelligence Lab., Univ. of Louisville, Louisville, KY 40292)

A short description of experiments that aim to determine visual cues influence on the perception of spatial sound is provided in the paper. The earlier stage of the carried out experiments showed that there exists a relationship between the perception of video presented in the screen and sound signals reproduced in a surround system. However, this relationship is dependent on the type of audio–visual signals. Thus a series of subjective tests has been performed on dozens of experts in order to discover these dependencies. The main issue in such experiments is the analysis of the influence of visual cues on the perception of the surround sound. This problem is solved with the application of genetic algorithm and rule searching mechanism to the processing of subjective test results. Some results and conclusions concerning the complexity of the investigated problem are included. [Research sponsored by the Committee for Scientific Research, Warsaw, Poland, Grant No. 8 T11D 00218.]

2pPP4. The "opposite direction shift" phenomenon in sound localization using two sound sources. Hiroyuki Fujii and Kazuhiko Kakehi (CIAIR, Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya-shi 464-8601, Japan)

It has been known that in a situation where tone and noise are presented at once, the perceived direction of the tone is shifted to the opposite direction of the noise. However, it is still not very clear which cues for sound localization are responsible for this ''opposite direction shift'' phenomenon. In this study, three factors of two sounds were investigated: temporal configuration, frequency overlapping, or level difference of two sounds. The experiment was based on a subjective sound localization task of the designated sound source among two sound sources. The target sound was a 1 kHz or a 500 Hz tone, and the distractor sound was pink or white noise or, notched noise. A target sound was presented for the cases where the distractor sound was at 25° left or 25° right of the loudspeaker array. The perceived direction of the target sound was compared. The ''opposite direction shift'' phenomenon was observed for the cases where the distractor was longer in duration or had a larger level than the target, or is overlapping with the target frequency. These results suggest that the important perceptual factor of the ''opposite direction shift'' phenomenon is related to the distractor's ''saliency.''

2pPP5. Interaural coherence and the lateralization of noise by interaural level differences. William M. Hartmann and Zachary A. Constan (Michigan State Univ., East Lansing, MI 48824)

Coherent noise, with an interaural cross correlation of unity, and incoherent noise, where signals to left and right ears are independent, sound very different. The former produces a compact lateralized image. The latter produces a diffuse image that fills the head. Nevertheless, listening experiments show that coherent and incoherent noise images seem to be equally affected by small interaural level differences (ILD). The ILD threshold for lateralizing incoherent noise is less than half a dB greater than the threshold for lateralizing coherent noise. In this sense, the human binaural system appears to behave like an ideal level meter, insensitive to the wave form and envelope fine structures that determine coherence. The small discrepancy (less than 0.5 dB) can be understood from a standard model of loudness perception—incorporating critical band filtering, half-wave rectification, amplitude compression (0.6 power law), and temporal integration (300 ms). The same model can also account for the small observed difference in lateralization ILD thresholds for coherent noise ($N0$) and antiphasic noise ($N\pi$). It can further account for the observed noise bandwidth dependence of lateralization ILD thresholds. [Work supported by the NIDCD.]

2pPP6. Detectability of phase change and its computational modeling. Masato Akagi and Masato Nishizawa (School of Information Sci., Japan Adv. Inst. Sci. & Tech., 1-1 Asahidai, Tatsunokuchi, Ishikawa 923-1292, Japan, akagi@jaist.ac.jp)

The ability to detect a phase change was measured in paired-stimuli comparison-experiments using monotonic phase (MPH) and phase-shifted (SPH) complex tones. Fundamental frequencies ($F0$) were 62.5 to 250 Hz in 1/4-octave steps. The number of harmonics was set to cover frequencies, 300–9300 Hz. All of the harmonics except one were started in sine phase. The remaining component was shifted pi (SPH) or not shifted (MPH). Subjects indicated whether paired stimuli were ''the same or not'' and reported whether phase-shifted harmonics appeared to ''pop-out.'' A computational model based on AIM (Patterson, 1995) was constructed and the experimental results were interpreted from autocorrelograms (ACGs) calculated by the model. The findings are as follows. When $F0$ was below 150 Hz and the number of harmonics in ERB at the phase-shifted harmonic exceeded 2, a phase change was detected and ''pop-out'' phenomena occurred. When $F0$ was over 200 Hz, however, MPH and SPH were not discriminated and there was no ''pop out.'' Discriminable paired stimuli have different ACGs and this reflects the detection ability of phase changes. Fine peaks in the ACG of SPH were found to correspond to the periods of the phase-shifted harmonics. [Work supported by CREST, JST.]

2pPP7. Backward interference of easily identified tones by simple and complex maskers. Richard E. Pastore and Melody S. Berens (Psych. Dept., Binghamton Univ., Binghamton, NY 13902, pastore@binghamton.edu)

There are several different types of interference in the perception of brief target tones due to the presence of subsequent stimulus in close temporal proximity. Backward masking is interference in the detection of the target tone, with the masker usually either a tone or noise. Backward recognition is interference of a tone masker in the discrimination of target tones typically separated by no more than a few jnd's. Components of natural temporal sequences are usually more complex than tones and more structured than noise, with major distinctions contrasted by more than a few jnd's. For example, brief initial position stop consonants are followed by complex vowel formants. The current study examines the ability to differentiate initial tones that, in isolation, are easily detected and distinguished (a musical fifth). In addition to noise, subsequent masking stimuli consisted of all possible combinations of three tones that are easily differentiated from each other (a musical third) and from the target tones. There is little backward interference from noise, small to moderate interference from individual tones, and significant interference from specific combinations of tone maskers. Thus backward interference is not limited to near threshold detection or discrimination.

2pPP8. Pitch interval perception by musically trained listeners. Henning Reetz (Linguist. Dept., Univ. of Konstanz, D186, 78462 Konstanz, Germany, henning.reetz@uni-konstanz.de)

This research investigates the reaction times to identify musical intervals by subjects with some musical training. All intervals within an octave, upward and downward, consisting of all combinations sine tones, complex tones with a fundamental, and complex tones without a fundamental were presented to a group of 18 subjects. Subjects had to decide which signal pairs form a specific musical interval. For some subjects the identification of certain intervals (e.g., a third major upward) was only 20 ms longer than the identification of two tones with identical pitch but the overall reaction time doubled when the same interval was presented downward. Some subjects showed a great dependency on the presence and absence of the fundamental whereas other subjects were not influenced by this parameter. Furthermore, the sort of musical instrument played by the subjects seem to influence the speed of perception of certain intervals. In total, reaction times indicate that the perception of the missing fundamental is idiosyncratic and depends on the musical education of the subjects and is not necessarily a part of normal auditory processing.

2pPP9. Effect of pitch fluctuation of vowels on vowel perception under a noisy environment. Kentaro Ishizuka and Kiyoaki Aikawa (NTT Commun. Sci. Labs., NTT Corp., Morinosato-Wakamiya 3-1, Atsugi City, Kanagawa 243-0198, Japan)

Previous paper reported that natural vowel identification scores were higher under harmonic complex tones than under a white noise, in spite of the tone and the noise having the same flat spectral envelope [K. Ishizuka and K. Aikawa, J. Acoust. Soc. Am. **109**, 2439 (2001)]. Natural vowels have pitch fluctuation. This paper reports a new finding that the pitch fluctuation has the effect of improving vowel identification under noisy environments. An experiment was conducted to examine the effect of pitch fluctuation of vowels on vowel perception under a harmonic complex tone at SNRs ranging from -10 to -15 dB. The Japanese female vowels were synthesized with the pitches fluctuated according to the $1/f$ characteristics [N. Aoki and T. Ifukube, J. Acoust. Soc. Am. **106**, 423–433 (1999)]. Five pitch fluctuation bandwidth conditions for vowels were tested. The bandwidths ranged from 0 to 4.3% ($F0 = 255.3$ Hz). The harmonic complex tone was composed of the same $F0$ of the vowels and 21 components in the cosine phase. The subjects were forced to choose the vowel they heard for each stimulus. The mean vowel identification scores increased from 33.2% to 46.5%, proportional to the increase of the pitch fluctuation bandwidth.

2pPP10. A method for the automatic hearing aid fitting employing speech in noise and fuzzy logic. Bozena Kostek (Inst. of Physiol. and Pathol. of Hearing, Pstrowskiego 1, 01-912 Warsaw, Poland) and Andrzej Czyzewski (Sound and Vision Eng. Dept., Tech. Univ. of Gdansk, Narutowicza 11/12, 80-952 Gdansk, Poland)

Some limitations of the hearing aid fitting process are discussed. The classical procedures in this process are based on audiometric test results and/or the loudness scaling method employing artificial test signals. However, the fitting of hearing aids should be also performed on the basis of testing speech understanding in noise, because this is much closer to the real life conditions. A satisfying reliability of these tests may be achieved through the use of modern computer technology with an application of a properly calibrated sound system. A new strategy applicable to fitting prostheses was developed. It allows finding automatically characteristics

of a hearing aid matching patients needs. The principles of the fitting method employing fuzzy reasoning, and some results of the experiments will be presented in the paper.

2pPP11. Computer simulations of hearing aid acoustical system performance. Grzegorz Szwoch (Sound and Vision Eng. Dept., Tech. Univ. of Gdansk, Narutowicza 11/12, 80-952 Gdansk, Poland) and Bozena Kostek (Inst. of Physiol. and Pathol. of Hearing, Pstrowskiego 1, 01-912 Warsaw, Poland)

The recent developments in the hearing aid technology enabled a number of improvements in hearing aids. This includes advanced signal processing algorithms, better speech intelligibility, miniaturization etc. One of the existing limitations is, however, the problem with providing patient-related characteristics of the acoustical system of a hearing aid. The aim of this paper is to show that using the physical modeling method it is possible to first build a model of the acoustical system of a hearing aid and then to simulate its performance. The waveguide model of the acoustical system of a hearing aid is proposed. Exemplary results of the computer simulations using such a model are presented and compared with some measurement data of existing hearing aid acoustical systems. The model proved to behave similarly to the real system. Conclusions regarding the application of such a method in the fitting process of a hearing aid are included.

2pPP12. Adaptive dual-microphone directional noise reduction scheme for hearing aids using nearby microphones. Martin Kompis and Matthias Bettler (Univ. Clinic of ENT, Head and Neck Surgery, Inselspital, 3010 Berne, Switzerland)

A new directional dual-microphone noise reduction scheme for hearing aids is presented. The acoustic field is sampled by two nearby microphones, which can be mounted conveniently in a single behind-the-ear or in-the-ear unit, such as a hearing aid or the external part of a cochlear implant system. A combination of time-invariant and adaptive postprocessing of the microphone signals is employed. First, time-invariant signal processing is used to combine the two microphone signals into two intermediate signals, one containing predominately sounds arriving from the front, one predominately from the rear. These signals are postprocessed by an adaptive noise canceller using filter lengths between 1 and 50 ms. Experiments have been performed in several simulated, realistic acoustic environment. In a moderately reverberant room (reverberation time 0.4 s, volume 34 m^3) improvements in signal to noise ratio (SNR) of 3.4–8.7 dB were achieved, depending on the length of the adaptive filter and the angle of incidence of the noise signal. A classic time-invariant dual-microphone approach yielded 0.2–6.9 dB lower SNRs in the same acoustic settings. The computational load of the proposed system is well within reach for current hearing aid technology. [Work supported by the Swiss National Science Foundation.]

2pPP13. Model studies of source term and reflection components of synchronous-evoked and distortion product otoacoustic emissions. Carrick Talmadge (Univ. of Mississippi, Natl. Ctr. for Physical Acoust., Oxford, MS 38677) and Arnold Tubis (Univ. of California, San Diego, La Jolla, CA 92093)

The unified model of synchronous-evoked, distortion product, and spontaneous otoacoustic emissions of Talmadge *et al.* [J. Acoust. Soc. Am. **104**, 1517 (1998); **105**, 275 (1999); **108**, 2570 (2000)] provides a useful testing ground for various paradigms that have been proposed for separately determining source terms and reflectance components of synchronous-evoked and distortion-product otoacoustic emissions. The source terms are the ear canal components that would exist in the absence of effects associated with cochlear wave reflections due to, e.g., distributed inhomogeneities [Zweig and Shera, J. Acoust. Soc. Am. **98**, 2018 (1995)] and nonlinearity [Talmadge *et al.*, 2000]. Model calculations are used to compare results for the unmixing of these components obtained using reflection-suppression techniques [see, e.g., Kalluri and Shera, J. Acoust. Soc. Am. **109**, 622 (2001)] and various forms of time-domain latency windowing ones [see, e.g., Knight and Kemp, J. Acoust. Soc. Am. **107**, 457 (2000)]. Because of multiple internal wave reflections at the cochlear base and the tonotopic place for the cochlear waves, the model implies (and experimental data indicate) that the reflectance components will generally exhibit fine structure effects similar to those of the total ear canal signal.

2pPP14. High-frequency stimulation in tinnitus treatment. Martin L. Lenhardt, Douglas G. Richards, Alan G. Madsen (Prog. in Biomed. Eng., Virginia Commonwealth Univ., Box 980158 MCV, Richmond, VA 23298-0168), Abraham Shulman, Barbara A. Goldstein, and Robert Guinta (Martha Entenmann Tinnitus Res. Ctr., Brooklyn, NY 11203)

Limited but convincing evidence suggests neural plasticity plays a role in the development and maintenance of tinnitus. Tinnitus is generally pitch matched with high frequencies (<6 kHz) and is often associated with very hearing loss, although the conventional audiometric thresholds may be normal or near normal. It is hypothesized that the brain experiences high-frequency air conduction auditory deprivation and this is a causative factor in tinnitus. If the low-pass middle ear is circumvented by bone conduction stimulation, residual high-frequency receptors in the cochlea can again be activated. High-frequency bone conduction transducers were fabricated to deliver frequencies from 6 to 40 kHz. The transducers were aluminum ceramic bimorphs with circular topology which were heavily damped by mass loading to the mastoid with resonances at 9 and 39 kHz. Ten subjects with persistent tinnitus were selected to listen to proprietary high-frequency patterned stimulation for two half-hour sessions for 4 weeks. Exit questionnaires revealed satisfaction with tinnitus relief in the majority (60%) with residual inhibition lasting on the order of weeks (mean=1.5 weeks). Encouraged by these findings, more extensive trials are underway to explore the efficacy of high-frequency bone conduction treatment in central tinnitus.

2pPP15. Digital USB audiometer with noise-shaping techniques for signal generation. Antonio Minguez, Juan Sancho, Jorge Grundman, and Vladimir Ulin (EUIT Telecomunicacion, Ctra. Valencia km. 7, 28031 Madrid, Spain)

A digital audiometer has been implemented on a personal computer with new features to make a low cost and simple equipment. The signals are generated digitally (digital samples) by a recursive algorithm that shapes the spectrum of the quantization noise of the digital samples in order to: (1) eliminate the harmonic distortion at low levels and (2) to increase the signal-to-noise ratio of every tone generated, as the standard EN 60645-1 requires for tone audiometers. Besides, the digital output of the personal computer (PC), the universal serial bus (USB), is used to link the PC with a compatible USB headphones so the transducer can get the original signal in a digital format, avoiding all the analog and electrical noise and distortion in the PC environment, so the proposed audiometer does not need any external equipment to generate the test signals, reducing the total cost of the product.

2pPP16. Implementation of narrow-band noise algorithms according to ANSI S3.6-1996. Mariano Ruiz, Ruben Salvador, and Manuel Recuero (Universidad Politecnica de Madrid, INSIA Crta. Valencia Km-7, Madrid 28031, Spain)

In order to generate narrow-band noise according to ANSI S3.6-1996 in an efficient way, it is necessary to employ multirate signal processing techniques. Thanks to the successive conversion of the sampling rate in the stages of the system, an efficient implementation of the required algorithms is obtained. These signals are entirely bandpass signals. The adopted solution is to design equivalent lowpass filters followed by a frequency modulation of the filtered signal with the purpose of moving it

to the desired frequency. To obtain a better performance in the resulting filter, multistage structures have been used to achieve the required changes in the sampling rate to implement the lowpass filter by means of a multirate signal processing technique. A LabVIEW application have been implemented for designing these types of filters. This program generates the filter coefficients and offers the possibility of carrying out a simulation and saving these coefficients into a C header file. These filters have been successfully tested, on a new DSP based audiometer that uses TI TMS320C6701, by means of Code Composer Studio and a series of routines that implement each one of the stages that compose the filter.

TUESDAY AFTERNOON, 4 DECEMBER 2001 ROOM 221, 2:00 TO 5:00 P.M.

Session 2pSA

Structural Acoustics and Vibration: Methods of Passive Dissipation in Structural Acoustics II

Courtney B. Burroughs, Chair
Applied Research Laboratory, Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

Invited Papers

2:00

2pSA1. Structural damping as a result of piezoelectric energy harvesting. George A. Lesieutre (Aerosp. Eng., Penn State Univ., University Park, PA 16802), Heath Hofmann, and Geffrey Ottman (Elec. Eng., Penn State Univ.)

This paper describes an approach to harvesting electrical energy from a mechanically-excited piezoelectric element. If the piezoelectric element is attached to a vibrating structure, this energy transfer also results in structural damping. The energy harvesting approach is loosely based on the fundamental electrical engineering concept of optimal loading. The harvesting circuit consists of an ac–dc rectifier with an output capacitor, an electrochemical battery, and a switch-mode dc–dc converter that controls the energy flow into the battery. An adaptive control technique was initially used to vary the switching duty cycle so as to maximize power transfer into the battery. Experiments demonstrated that the adaptive dc–dc converter increased power transfer by over 400%, as compared to direct connection of the rectifier output to the battery. For persistent excitation above a certain level, the existence of a near-optimum duty cycle was also observed. A standalone energy harvesting system that exploited this observation was developed. This system has two modes: for low excitation, the rectifier charges the battery directly; while for higher excitation, the battery runs a fixed duty cycle dc–dc converter. This standalone system was successfully demonstrated in the laboratory. Effective damping loss factors associated with this energy harvesting scheme are also reported.

2:30

2pSA2. Real-time switching of piezoelectric shunts for structural vibration control. William W. Clark (648 Benedum Hall, Mech. Eng. Dept., Univ. of Pittsburgh, Pittsburgh, PA 15261)

The use of piezoelectric actuators to passively add damping to vibrating structures has been investigated for over a decade. Simple resonant shunt circuits consisting of resistors and inductors are connected to the piezoelectric actuators to add damping to one of the structural modes [Hagood and Crawley, J. Guid. Control Dyn. **14**]. Modifications to the method have been proposed to enable a multi-modal damping technique, but at the cost of performance to any one mode [e.g., Wu, Proc. of 1998 SPIE Smart Structures and Materials Conference, Vol. 3327, pp. 159–168]. The objective of this talk is to discuss recent research into methods that help eliminate the disadvantages of passive resonant shunt damping. The methods to be discussed involve real-time switching of piezoelectric shunt circuits, and are based on the foundations laid by Larson [Ph.D. Dissertation, Georgia Institute of Technology, 1996], Clark [J. Intell. Mater. Syst. Struct. **11**, 263–271], and Richard *et al.* [Proc. of 1999 SPIE Smart Structures and Materials Conference, Vol. 3672, pp. 104–111]. This talk will begin with the basic concepts of real-time switching, and will lead up to recent research that has demonstrated the use of the technique for multi-modal vibration suppression.

3:00

2pSA3. A nonlinear semipassive vibration damping using piezoelectric elements. Daniel J. Guyomar, Claude L. Richard, and David S. Audigier (INSA, LGEF, Bt 504, 8 Rue de la Physique, 69621 Villeurbanne, France)

The proposed approach is based on a nonlinear filtering of the piezoelements output voltage and does not require any tuning to ensure damping performances over a large bandwidth. The output voltage is short circuited for a very brief period of time when the voltage reaches a maximum or a minimum. As a consequence of this filtering, the voltage gets distorted, amplified, and shifted in the time domain. It can be split in two signals, one is in phase with the vibration displacement and contributes to the potential energy while the second is in phase with the vibration speed and participates to the energy damping. The technique is semipassive since it requires a very low electrical power to drive the switch sequences. Experimental data show that the switching technique leads to better damping performances than purely passive methods whatever the natural damping of the structural material is. Furthermore, switching the piezovoltage through a small inductor leads to a voltage amplification and consequently to an increase of the damping. Damping capabilities range from 6 dB for an epoxy beam to nearly 20 dB for a steel beam even in the low frequency regime. Experimental results and theoretical interpretations will be presented.

3:30

2pSA4. Absorption and trapping of vibration energy by complex resonators. Antonio Carcaterra (INSEAN, Istituto Nazionale per Studi ed Esperienze di Architettura Navale, Via di Vallerano, 139, 0128, Rome, Italy) and Adnan Akay (Mech. Eng. Dept., Carnegie Mellon Univ., Pittsburgh, PA 15143)

This presentation examines conditions that can enhance apparent damping of a master structure due to its coupling with a complex resonator that does not exhibit classical dissipation properties. The model uses a set of multistage parallel oscillators within which energy is trapped. The method presented here for predicting the apparent damping is based on evaluation of the asymptotic response of the master structure utilizing the Hilbert transform. The result is a functional relationship between the frequency distribution within the system and the resulting apparent damping. This study leads to a nonclassical dynamic absorber, consisting of a multistage parallel oscillators, in terms of its absorption and trapping characteristics. The energy is absorbed from the master at a rate determined by the apparent damping and trapped for a given time interval as measured by the return time. The theoretical results suggest optimal configurations to decrease the absorption time and to increase the trapping time (up to an infinite delay). The results of numerical simulations performed on complex vibro-acoustic systems support the proposed theoretical approach. [Research sponsored by INSEAN and NSF.]

4:00–4:15 Break

Contributed Papers

4:15

2pSA5. Use of the two channels method to determine the guided waves damping coefficients of an asymmetrically fluid-loaded plate: Theory and experiment. Ech-Cherif El-Kettani Mounsif and Franklin Herve (Laboratoire d'Acoustique Ultrasonore et d'Electronique, L.A.U.E. U.M.R. C.N.R.S. 6068, Universit du Havre, Pl. Robert Schuman, 76610 Le Havre, France)

Exact and approximate formalisms describing the interactions of acoustic plane waves with an elastic isotropic plate immersed between two different fluids are presented. This is an extention of the Fiorito, Madigosky, and Berall (FMU) theory. Resonant approximations are given for the reflection and transmission coefficients, which assume light fluids loading. These approximations show that the resonance widths are the sum of two independent partial widths, each of them being related to one fluid and to the plate physical properties. In addition, experimental results are presented for a steel plate; the damping coefficients of the propagating leaky Lamb waves along the loaded plate are measured in two different loading situations. In the first situation the plate is embedded in water (water loading on both sides), while in the second situation it is loaded with water on one side and acetone on the other side. The experimental damping coefficient of the acetone-steel-acetone structure is deduced therefrom. The damping coefficient obtained in each case is then compared to the theoretical half resonance width, and the experimental results are found to be in good agreement with the theoretical predictions.

4:30

2pSA6. Analytical solutions for characterizing tortuosity and characteristic lengths of porous materials using acoustical measurements: Indirect model. Xavier Olny, Jerome Tran Van (Ecole Nationale des Travaux Publics de l'Etat, DGCB URA CNRS 1652, rue Maurice Audin, 69518 Vaulx en Velin, France), and Raymond Panneton (GAUS, Dept. of Mech. Eng., Univ. de Sherbrooke, QC J1K 2R1, Canada)

Under the homogenization hypothesis, wave propagation in porous materials can be described using, for example, macroscopic dynamic density and bulk modulus. These functions depend on statistical geometrical parameters. Classically, five parameters are used: porosity, static air flow resistance, tortuosity, and viscous and thermal characteristic lengths. The three last parameters are generally difficult to measure with existing direct methods, for a wide range of materials. The proposed method is based on the measurement of dynamic density and compressibility, in order to separate viscous and thermal effects. With prior knowledge of airflow resistance and porosity, it is then possible to find analytical solutions for the missing parameters, fitting Johnson–Allard–Champoux model. In the same way, we also tackle the problem of the determination of static ''thermal permeability'' introduced to improve the description of thermal dissipation effect. Experimental results, obtained with various materials, using a Kundt's tube, are presented to show the efficiency, and relative simplicity of the method. Moreover, the relevance of using a three parameters model for describing the bulk modulus is discussed.

4:45

2pSA7. Analytical solutions for characterizing tortuosity and characteristic lengths of porous materials using acoustical measurements: Extrapolation model. Raymond Panneton (GAUS, Dept. of Mech. Eng., Univ. de Sherbrooke, QC J1K 2R1, Canada), Xavier Olny, and Jerome Tran Van (Ecole Nationale des Travaux Publics de l'Etat, DGCB URA CNRS 1652, rue Maurice Audin, 69518 Vaulx en Velin, France)

Following sophisticated descriptive models on sound propagation in porous media, five acoustical parameters usually describe the complexity of the porous network: porosity, static air flow resistance, tortuosity, and viscous and thermal characteristic lengths. While porosity and airflow resistance can be measured directly and accurately using simple apparatuses, simple and accurate direct measurements of the tortuosity and characteristic lengths are still an open research area. Actual researches on the characterization of these parameters use acoustical methods in the ultrasound frequency range. In this paper, a similar approach is investigated, but applying this time in the frequency range 2000–6000 Hz. The method is based on analytical solutions linking the tortuosity and characteristic lengths to measured dynamic density and compressibility. Following an extrapolation procedure to infinite frequency, the tortuosity and characteristic lengths are deduced without prior knowledge of porosity and air flow resistance. Experimental results from Kundt's tube are compared to simulations on various materials to show the accuracy and limitations of the method.

Session 2pSC

Speech Communication: Cross-Language Issues (Poster Session)

Megha Sundara, Chair
School of Communication Sciences and Disorders, McGill University, 1266 Pine Avenue, West, Montreal, Quebec H3Q 1A8, Canada

Contributed Papers

All posters will be on display from 1:30 p.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:15 p.m. and contributors of even-numbered papers will be at their posters from 3:15 p.m. to 5:00 p.m.

2pSC1. On talker voice in language identification. Verna Stockmal and Z. S. Bond (Dept. of Linguist., Ohio Univ., Athens, OH 45701)

Listeners find that discriminating between two languages is relatively easy compared to identifying new samples. Listener responses seem to be influenced by regional speech characteristics and talker voice quality. This study attempted to assess listener ability to match spoken samples of unknown languages when produced by male and female talkers. Samples from Arabic and Latvian, three provided by male talkers and three by female talkers, were assembled in a test using ABX format. Listeners matched the X language with either the A or B language, across same sex and different sex talkers. Overall, listeners performed at above chance levels, 60% correct. Listeners identified the two languages at approximately equal rates. They matched speech samples produced by males (75%) better than speech samples produced by females (53%). Listeners matched languages across speaker gender at 55% correct. These results suggest that speaker sex, as part of voice quality, is encoded with a representation of the unknown language.

2pSC2. Non-native speech production I: Listener–talker adaptation. Ann R. Bradlow and Tessa Bent (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208)

To assess native listener adaptation to foreign-accented speech, we submitted 16 sentences produced by three talkers of Chinese-accented English to intelligibility testing. First, the test sentences were orthographically transcribed by native English listeners in a single-talker format in which the test sentences from one talker were preceded by 32, and followed by 16, other sentences produced by the same talker. Second, the test sentences were presented to native listeners in a multiple-talker format in which the test sentences from one talker were interspersed with 48 similar sentences produced by 31 other non-native talkers from various language backgrounds. Within the single-talker condition, the native listeners' transcription accuracies for all three talkers improved from the beginning to the end of the test session. The magnitude of this improvement was similar across all three talkers. Furthermore, for all three talkers, intelligibility was better in the single-talker condition than in the multiple-talker condition; however, this difference depended on the talker's overall intelligibility. The ''best'' and ''worst'' talkers showed smaller single- versus multiple-talker differences than the ''medium'' talker. These findings demonstrate that foreign-accented speech intelligibility can improve as the listener gains experience with the voice and speech patterns of the individual non-native talker. [Work supported by NIH-NIDCD Grant No. DC03176.]

2pSC3. Non-native speech production (II): phonemic errors by position-in-word and intelligibility. Tessa Bent (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, t-bent@northwestern.edu), Bruce L. Smith, Danielle Lodewyck (Dept. of Commun. Sci. and Disord., 2299 N. Campus Dr.), and Ann R. Bradlow (Dept. of Linguist.)

This study considered the distribution of non-native talker phonemic errors by position-in-word, and their differential effects on non-native speech intelligibility. Based on the structure of words in the native language of the talkers (i.e., Chinese), it was hypothesized that word-initial phonemes would be produced more accurately than word-final phonemes. Additionally, based on models of lexical access (e.g., cohort model), it was hypothesized that phonemic errors in the word-initial position would be more detrimental to intelligibility than phonemic errors later in the word. 15 talkers of Chinese-accented English produced 16 simple declarative sentences. These recordings were phonetically transcribed to determine the number and type of production errors and presented to native English listeners for intelligibility testing. Results showed that, as predicted, word-initial phonemes were produced more accurately (92% correct) than word-final phonemes (82% correct) [$t(14)=3.815$, $p<0.01$]. Furthermore, word-initial singleton consonants and vowels correlated with intelligibility (Rho$=0.738$, $p<0.01$ and Rho$=0.692$, $p<0.01$, respectively). In contrast, initial and final consonant clusters, word-final singleton consonants, and within-word consonants were not significantly correlated with intelligibility. These results demonstrate that, for a variety of talker- and listener-related reasons, a phoneme's position within the word exerts an important influence on both its production and perception. [Work supported by NIH-NIDCD Grant No. DC03176.]

2pSC4. The pronunciation of English sentences by Korean children and adults. J. E. Flege (Div. of Speech & Hearing Sci., Univ. of Alabama at Birmingham, CH20, 1530 3rd Ave. S., Birmingham, AL 35294, jeflege@uab.edu), David Birdsong (Univ. of Texas, Austin, TX 78712), Ellen Bialystok (York Univ., Tononto, ON M3J 1P3, Canada), Molly Mack (Univ. of Illinois, Urbana, IL 61801), and Hyekyung Sung (Stanford Univ., Stanford, CA 94305)

This study examined English sentences produced by four groups of native Korean subjects (18 each) who differed according to mean age (children$=12$ years, adults$=32$ years) and length of residence in North America (means$=3$ vs 5 years). A delayed repetition technique was used to elicit English sentences at Time 1 and one year later at Time 2. Native English-speaking listeners used a 9-point scale to rate the sentences for overall degree of foreign accent. The ratings obtained for the native Korean (NK) subjects were converted to z-scores using the mean ratings and standard deviations obtained for sentences produced by control groups of Native children and adults. As expected, analyses of the standardized rat-

ings revealed that the NK children produced the sentences with milder foreign accents than the NK adults did at both Time 1 and Time 2. Unexpectedly, the adult–child difference was larger at Time 2 than Time 1 because the NK children's foreign accents diminished whereas the NK adults' foreign accents grew significantly stronger from Time 1 to Time 2. Possible explanations for this are age-related differences in motivation, English input, or the strength of influence of Korean phonetic structures on the English sound system.

2pSC5. Intonation patterns in productions by Seoul Korean speakers of English. Julia Tevis McGory (Ohio State Univ., Dept. of Speech & Hearing, 110 Pressey Hall, 1070 Carmack Rd., Columbus, OH 43210)

This study compares the intonation patterns produced in English declaratives and yes–no questions by native American English speakers and native Korean speakers. As is well known, stress and pitch accent are integral to English intonation. Declaratives are produced with an *F*0-rise (L+H*) in the stressed syllable of the most prominent word followed by a steep utterance-final fall while yes–no questions are produced with a sharp *F*0-fall (L*) onto the stressed syllable followed by a gradual utterance-final rise. Words that follow the focused word are not associated with any particular *F*0 pattern. In contrast, Korean has neither stress nor pitch accent. Prominent words in statements and questions are produced at the beginning of a phrase and realized with an *F*0-rise (LH) followed by a gradual fall to the end of statements (L%) and a final rise (LH%) in questions. In this study, subjects from both subject groups produced the words ''memorial'' and ''memorizes'' in both focus and post-focus positions in both statements and yes–no questions. The question of interest is whether Korean speakers follow the intonational pattern of English (their L2) or utilize the intonational characteristics of their native language.

2pSC6. Bilingual speech production: The case of simultaneous acquisition. Megha Sundara and Linda Polka (School of Commun. Sci. and Disord., McGill Univ., 1266 Pine Ave. W., Montreal, QC H3G 1A8, Canada)

Despite increased interest in people learning a second language (L2) in the past decade or two, there have been few studies investigating production by bilinguals at the segmental level. In this experiment coronal stop consonants in real-word productions in French and English by bilingual French–English speakers, very fluent in both languages having learned the two languages simultaneously since birth, were recorded. Differences in the phonetics of Canadian French and English include VOT and place of articulation distinctions. Canadian French has a two-way voicing distinction between prevoiced and simultaneous release stops and uses the dental place of articulation. Canadian English, on the other hand, has a two-way voicing distinction (in syllable initial position) between simultaneous and aspirated release for stops and uses an alveolar place of articulation. Acoustic analyses of stop consonants for both VOT and place of articulation (dental versus alveolar) are reported. Productions by bilinguals will be compared with those of respective monolingual groups for similarities and differences. Results will be discussed with respect to Grosjean's wholistic-fractionated classification of the bilingual system [F. Grosjean, Brain and Language **36**, 3–15 (1989)].

2pSC7. Articulatory gestures of /l/ by bilingual Korean–English children. Sunyoung Oh (Dept. of Linguist., Univ. of British Columbia, E270-1866 Main Mall, Vancouver, BC V6T 1Z1, Canada)

Properties of articulatory gestures (e.g., timing and magnitude) have been used to find the phonetic correlates of syllable position. For example, allophones of English /l/ are determined by the timing relationship between the tongue tip and dorsum gestures, where greater timing lag is observed in the post-vocalic position. On the other hand, allophones of Korean /l/ are determined by the position of the tongue tip, prevocalic post-alveolar versus post-vocalic alveolar. The magnitude of both gestures is also greater in post-vocalic position, and the tongue dorsum is slightly advanced in Korean. This paper examines cross-language influences of the articulatory gestures of /l/. To test this, an ultrasound study of the speech of two bilingual Korean–English children was conducted. The results showed that bilingual children produced /l/ using the correct tongue tip gestures of Korean in both positions (post-alveolar versus alveolar); however, they failed to implement tongue dorsum advancement. Instead, they incorrectly adopted the English-like gesture and implemented tongue dorsum retraction in post-vocalic positions. These indicate that bilingual children are attuned to language-specific gestures of both languages and there is at least a one-way effect (English to Korean) on /l/ gestures.

2pSC8. Perception and production of English consonants by Japanese adults and children. Katsura Aoyama (Div. of Speech & Hearing Sci., Univ. of Alabama at Birmingham, CH20, 1530 3rd Ave. S., Birmingham, AL 35294, aoyama@uab.edu), Susan Guion (Univ. of Oregon, Eugene, OR 97403), J. E. Flege (Univ. of Alabama at Birmingham, Birmingham, AL 35294), Tsuneo Yamada (National Inst. of Multimedia Education, Chiba 261-0014, Japan), and Reiko Akahane-Yamada (ATR Human Information Processing Res. Labs., Kyoto 619-0288, Japan)

This study examined perception and production of English consonants by native Japanese (NJ) adults and children (16 subjects in each group, mean age=39.9 and 9.9 years). The subjects were tested 2 times (T1, T2), when their mean lengths of residence in the U.S. were 0.5 years and 1.6 years. Age-matched native English (NE) controls (16 adults, 16 children) also participated. A categorial discrimination task was used to assess the perception of the English contrasts /b/–/s/, /b/–/v/, /ɹ/-/l/, /ɹ/–/w/, and /s/–/θ/. For three groups (NE adults, NE children, NJ adults), the T1 and T2 scores did not differ significantly on any contrast. The NJ children's scores were significantly higher at T2 than at T1 on the /ɹ/–/l/ and /ɹ/–/w/ contrasts. A picture naming task with auditory cues were used to elicit the production of English words containing consonants of interest. Twelve NE listeners identified the subjects' production of /ɹ/, /l/, and /w/ in word-initial position. Only the NJ children's scores improved significantly from T1 to T2, and their improvement was the greatest on the production of /ɹ/. Several possible explanations will be offered for why there seems to be more improvement from T1 to T2 for /ɹ/ than /l/.

2pSC9. Acoustic analysis of American listeners' Mandarin tone productions before and after perceptual training. Yue Wang (Ctr. for Mind, Brain, and Learning, Box 357988, Univ. of Washington, Seattle, WA 98195), Allard Jongman, and Joan A. Sereno (Linguist. Dept., Blake Hall, Univ. of Kansas, Lawrence, KS 66045)

Training American listeners to perceive Mandarin tones has been shown to be effective, with the trainees' identification improving by 21% [Y. Wang, A. Jongman, and J. Sereno, J. Acoust. Soc. Am. **106**, 3649–3658 (1999)]. The effect on production as a result of perceptual training was first assessed by native Mandarin listeners' evaluation of the trainees' tone productions, showing that post-training productions were 18% more accurately identified than pre-training productions [Y. Wang, M. Spence, A. Jongman, and J. Sereno, J. Acoust. Soc. Am. **105**, 1095 (1999)]. A detailed acoustic analysis has now been performed to further examine the nature of the production improvement. Trainees' productions include 80 Mandarin words recorded both before and after perceptual training. For comparison, these words were also produced by native Mandarin speakers. For each production, a pitch track was derived. Results show that post-training tone contours approximate native norms to a greater degree than pre-training tone contours, indicating an improvement in production following perceptual training. Furthermore, pitch height and pitch contour are not mastered in parallel, with the former being more resistant to improvement than the latter. The results are discussed in terms of the production–perception relationship and native language influence.

2pSC10. Discrimination of Thai tones by naive and experienced native English speakers. Ratree P. Wayland and Susan G. Guion (Dept. of Linguist., 4131 Turlington, Univ. of Florida, Gainesville, FL 32611-5454)

This study investigated the ability to discriminate between the low and the mid tones of Thai by two groups of native English speakers: naive and experienced. The naive group was comprised of eight native American English speakers who had no previous exposure to Thai or any tonal language. The eight native English speakers in the experienced group had been learning and had lived in Thailand for a varying amount of time. Eight native speakers of Thai participated as control subjects. Stimuli were eight minimal pairs. Three variables were manipulated: (1) experience with the target language: naive vs experienced; (2) syllable types: closed vs open; and (3) levels of processing: phonetic vs phonological implemented by using two different ISI (500 ms and 1500 ms). It was found that experience with the target language resulted in improved discrimination of the target language's tonal contrast. Moreover, all three groups of subjects performed better with closed syllables than open syllables. Contrary to previous research, the effect of ISI was found only for the most difficult contrasts and only for the experienced group. An explanation for the findings and implication for futher research will be offered and discussed.

2pSC11. Perception of English vowels by native Korean and Japanese speakers. Takeshi Nozawa (Kansai Univ. of International Studies, 1-18 Aoyama Miki, Hyogo 673-0521, Japan) and Elaina M. Frieda (Dept. of Speech and Hearing Sci., Ohio State Univ., 110 Pressey Hall, 1070 Carmack Rd., Columbus, OH 43210-1002)

Native speakers of Korean and Japanese were assessed in their ability to perceive English vowels categorically. These two languages were chosen because they differ in the size of their vowel inventories with Korean containing more vowels than Japanese. The Korean vowel space is presumably partitioned more so than the Japanese and this may facilitate Korean speakers' discrimination of English. This hypothesis is in keeping with Best's Perceptual Assimilation Model (PAM). PAM states that non-native speakers' discrimination of non-native sounds can be predicted by differing degrees based on whether the new non-native sounds are assimilated to one or more native phonemic categories. Further, discrimination is mediated by how well the non-native category fits the native category. Both subject groups completed the following tasks: the Categorical Discrimination Test (CDT, designed by J. Flege) which tested their discrimination of English vowels; and an assimilation task where subjects equated English vowels to their native vowels and rated each for category goodness. A' scores were calculated for each contrast in the CDT. In addition, confusion matrices were calculated for each group. For both subject groups, the English vowels assimilated to a single vowel category were found to be more difficult to discriminate.

2pSC12. Perception of coarticulation with Japanese voiced and voiceless vowels. Tomomi Hashimoto (Prog. in Linguist., Univ. of Florida, 4131 Turlington Hall, Gainesville, FL 32611, tomomi@ufl.edu)

Two experiments investigate whether coarticulatory information of voiced and devoiced vowels in their preceding segments can be exploited by speakers of different dialects and/or by second language learners. In standard Tokyo Japanese, the high vowels [i, u] are devoiced between voiceless segments while vowel devoicing is not typical in Osaka dialect. This study examines how vowel's voicing status affects listeners' ability to exploit coarticulatory information in initial consonants. In both experiments, listeners consisted of native speakers of Tokyo and Osaka dialects and native speakers of English learning Japanese as a second language. In the first experiment, the post-alveolar fricative /(esh)/ followed by a vowel, either /i/ or /u/, was produced in real words by native speakers of Tokyo and Osaka dialects. The listeners were presented with a series of excised /(esh)/ to determine the original vowel identity. The second experiment replicated the first experiment testing with the velar stop /k/. The preliminary results suggest vowel devoicing may cause perceptual difficulties in the fricative-vowel context, but not in the stop-vowel context. The results also indicate a difference in cue weighting among 3 groups. Explanation for this finding will also be explored by way of acoustic measurements.

2pSC13. Characteristics and intelligibility of Japanese-accented English. Kanae Nishi (Univ. of South Florida, 4202 E. Flower Ave., Tampa, FL 33620)

The relationship between intelligibility and the degree of accentedness of Japanese-accented English was studied using an accentedness rating task, an identification task, and multidimensional scaling (MDS). Native speakers of American English (AE) listened to speech samples of 270 AE monosyllable words recorded by three groups of speakers (native AE, Japanese residing in the United States, and Japanese who had never lived outside Japan). The length of stay in the U.S. was positively related to native-like dimensionality in MDS solutions and the correct identification of words, but unrelated to ratings of foreign accentedness. High intelligibility was not necessarily achieved by less-accented speakers but by speakers with well-differentiated dimensional structures.

2pSC14. Native and non-native perception of phonemic length contrasts in Japanese. Hiroaki Kato, Keiichi Tajima, and Reiko Akahane-Yamada (Information Sci. Div., ATR Inst. Internal, 2-2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan, kato@isd.atr.co.jp)

Segment length is distinctive in Japanese, e.g., /kado/ ''corner'' versus /ka:do/ ''card,'' unlike languages such as English. To study this property extensively and precisely, a series of perception experiments were conducted. Stimuli were pairs of unaccented 3- or 4-mora nonsense Japanese words of the basic form /ere__e/ where __ is one of /p t k m n s/. Each pair contrasted in either vowel length, e.g., /erete/ versus /ere:te/, or in consonant length, e.g., /erete/ versus /erette/. Each pair was also embedded within a carrier sentence. A synthetic continuum gradually varying in the duration of the target vowel or consonant was built for each pair using STRAIGHT, a high-fidelity speech analysis, synthesis, and manipulation system [Kawahara *et al.*, Speech Commun. **27**, 187–207 (1999)]. Native Japanese and native English speakers participated in both identification and discrimination tests of each continuum. Preliminary results from stimuli containing stops indicate that the non-native discrimination scores were significantly poorer than the native scores, particularly when the continua were embedded in a sentence. The non-native identification scores, however, were similar to those of the native. Implications of these findings are discussed along with results from the fricative and nasal continua.

2pSC15. Cross-dialect perception of English vowels by native Japanese speakers. Elaina M. Frieda and Robert A. Fox (Speech Percept. and Acoust. Lab, Speech and Hearing Sci., Ohio State Univ., 110 Pressey Hall, 1070 Carmack Rd., Columbus, OH 43210-1002, frieda.1@osu.edu)

The present study will examine the influence of dialect on second language acquisition in terms of perceptual abilities. Our lab is currently comparing vowel productions from native Japanese (NJ) speakers with native American English speakers residing in two divergent dialectal regions of the United States (Alabama and Ohio). This study is an elaboration of this work that explores the extent to which there is regional variation in the perceptual responses in NJ as a function of local dialect. We recorded sets of words from four groups of subjects: NJ subjects residing in Alabama (JA) and Ohio (JO) and native English (NE) subjects from Alabama (EA) and Ohio (EO). These tokens were then used in an identification experiment utilizing the same four subject groups. The experiment addresses at least two different questions: (1) Does the intelligibility of Japanese vowels change as a function of dialect for the English listeners?

(2) Does the perceptual responses of the Japanese listeners to the EA and EO tokens vary as a function of their ambient linguistic context? The results will be discussed in terms of second language acquisition and dialect acquisition. [Work supported, in part, by an INRS Award from the NIH to Fox, P.I.]

2pSC16. Brain plasticity in behavioral and neuromagnetic measures: A perceptual training study. Yang Zhang, Patricia K. Kuhl (Ctr. for Mind, Brain, and Learning & Dept. of Speech and Hearing Sci., Univ. of Washington, Seattle, WA 98195), Toshiaki Imada (Nippon Telegraph and Telephone Corp., Atsugi, Kanagawa 243-0198, Japan), Makoto Kotani (Tokyo Denki Univ., Chiyoda-ku, Tokyo 101-8457, Japan), and John Pruitt (Microsoft Corp., Redmond, WA 98052)

The present study investigated brain plasticity in Japanese adults learning the English /l–r/ distinction. Both behavioral and brain (MEG) measures were used. Training was designed to mimic the listening experience of infants who are exposed to the exaggerated acoustic events contained in infant-directed speech (''motherese''). A multimedia software program incorporated major features conducive to perceptual learning. The program consisted of 12 1-h training sessions. Pre- and post-training measures were collected from nine subjects including two controls. The MEG experiments examined training effects in preattentive discrimination versus categorization conditions. An average improvement of 21.7% was obtained behaviorally. The training effect was significant for all factors of speaker, vowel, and syllabic context. Perceptual learning generalized to three untrained voices. No effects were observed in the two control subjects. Consistent with the behavioral results, the MEG data showed significantly improved mismatch responses in the left hemisphere. Both behavioral and brain data indicated that the learned phonetic boundary in Japanese subjects did not match that of the American listeners. This pattern of results suggests the existence of substantial neural plasticity in adulthood as well as a potential limit. [Work supported by NIH, HFSP, and the Talaris Research Institute.]

2pSC17. Effects of short-term exposure to a foreign language on discrimination of a non-native phonetic contrast: Convergent evidence from brain and behavioral tests. Patricia K. Kuhl, Feng-Ming Tsao, Huei-Mei Liu, Sharon Corina, Erica Stevens, Tobey Nelson, Jessica Pruitt, and Denise Padden (Ctr. for Mind, Brain, and Learning, Univ. of Washington, Seattle, WA 98195)

Studies in our laboratory demonstrate that between 6 and 12 months of age infants show a significant increase in the ability to discriminate native-language phonetic contrasts and a decline in foreign-language discrimination. The increase in performance on native-language contrasts suggests a process of active learning rather than maintenance. In the present experiment, we tested whether the learning process infants engage in during the period between 6 and 12 months extends to a foreign language the infants had not previously heard. American infants at 9 months of age participated in a 12-session language play group in which they heard a native speaker of Mandarin Chinese read, play, and talk to them. A control group was exposed to American English using the same books and toys. Both groups were subsequently tested on a Mandarin Chinese contrast using both behavioral and brain measures. The results demonstrate that infants exposed to Mandarin over a 4-week period show significant discrimination of the Mandarin contrast measured both behaviorally and with event-related potentials (ERP). The data, demonstrating that short-term exposure to language results in significant increases in performance, suggest a powerful learning process in language acquisition. [Research supported by NIH, HFSP, and the Talaris Research Institute.]

2pSC18. Developmental changes in the discrimination of Hindi dental-retroflex consonants by American–English children between childhood and puberty. Jesica C. Pruitt (Ctr. for Mind, Brain, and Learning & Speech and Hearing Sci., Univ. of Washington, Seattle, WA), John S. Pruitt (Microsoft Corp., Redmond, WA), Tobey L. Nelson, and Patricia K. Kuhl (Ctr. for Mind, Brain, and Learning & Speech and Hearing Sci., Univ. of Washington, Seattle, WA)

There are many studies comparing adult and infant perception of non-native speech sounds, with results typically showing that infants outperform adults on non-native phonetic discrimination. However, there is a lack of research that investigates the developmental change in native and non-native speech perception, particularly from early childhood to puberty, when plasticity for learning a new language appears to decline. In an attempt to better understand the effects of language experience on native and non-native speech perception during this period, the current study examined 6 to 15-year-old American–English children's ability to discriminate Hindi dental-retroflex consonants. Discrimination data were obtained using four phonetically different consonant–vowel contrasts that are phonemic in Hindi but not in English. Discrimination data were also obtained using a native-language contrast (AE /r/ vs /l/). For comparison, American–English and Hindi adults were tested. A newly developed computer-based testing method was used with all subject groups to present the various contrasts in an engaging and efficient manner. Findings suggest that there are interesting developmental trends in the age range of 6–15 years, and that the computer-based testing method will provide a valuable tool for assessing speech perception across a wide age range. [Work supported by NICHD and Talaris Research Institute.]

2pSC19. Word segmentation in a foreign language: Further evidence for crosslinguistic strategies. Barbara Hoehle, Dorothee Giesecke (Dept. of Linguist., Univ. of Potsdam, P.O. Box 60 15 53, 14415 Potsdam, Germany), and Peter W. Jusczyk (Dept. of Psych., Ames Hall, Johns Hopkins Univ., Baltimore, MD 21218-2686)

We report further evidence that early word segmentation is triggered by rhythmic properties of the language. Our study used the head turn preference paradigm to compare the skills of English and German 9-month-olds to extract bisyllabic trochaic German words from German text passages. Both English and German are stress-timed languages and have trochaic feet as their dominant stress pattern in bisyllabic words. Consequently, no differences in the segmentation strategies used by infants from these two language groups were expected. In line with our predictions, even though the stimulus material contained phonemes that do not belong to the English phoneme inventory and violations of English phonotactics, English-learners were as successful in detecting the words as German infants. These findings replicate those of Houston *et al.* (2000) with Dutch and English. Thus, our findings provide further evidence that children can exploit processing strategies developed in their own language to process another language. Furthermore, the data suggest that, at least initially, word segmentation in infancy relies on acoustical information in a pure bottom-up fashion without the support of existing knowledge about the language to be processed.

2pSC20. Eye movements and the activation of native words in non-native spoken-word recognition. Andrea Weber (MPI f. Psycholing., Wundtlaan 1, 6525 XD Nijmegen, The Netherlands and Speech and Hearing Dept., Grad. Ctr., CUNY, 365 5th Ave., New York, NY 10016-4309)

Using the eye-tracking paradigm, this study explored the recognition of non-native spoken words. In particular, it focused on whether listeners activate native candidate words during the recognition of non-native words. Dutch participants followed spoken instructions in English to click on pictures using a computer mouse, while their eye movements were monitored. A target picture (e.g., the picture of a desk) was presented along with distractor pictures. The Dutch name of a distractor picture was either phonologically related to the English name of the target picture

(e.g., English target *desk* /dEsk/ and Dutch competitor *deksel*, ''lid,'' /dEks@l/) or it was phonologically unrelated (e.g., *bloem*, ''flower'' or *schommel*, ''swing''). Eye movements to a picture were interpreted as evidence for activation of the lexical representation associated with this picture. As the English target word unfolded over time, participants fixated distractor pictures with phonologically related Dutch names more than distractor pictures with phonologically unrelated names. The results demonstrate that listeners consider candidate words of the native language for recognition of a non-native word. Listeners are not able to deactivate the native mental lexicon even when they are in a monolingual non-native situation where the native language is irrelevant and possibly even misleading. [Work supported by MPI for Psycholinguistics.]

Session 2pUW

Underwater Acoustics and Archives and History: How Did We Get Here? Insights into the History of Underwater Acoustics

Ralph R. Goodman, Chair
Applied Research Laboratories, Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804-0030

Chair's Introduction—12:55

Invited Papers

1:00

2pUW1. Project Artemis—A retrospective view. Alan Berman (Unit A-3, 2300 W. Alameda St., Santa Fe, NM 87501)

Approximately 45 years ago, the Underwater Acoustics Community of the U.S. undertook a massive research program called Project Artemis. The objective of this endeavor was to determine what were the necessary capabilities needed to build an active mono-static sonar that was capable of detecting a submerged submarine at ranges of about 1000 km. Although no operational capability ever resulted from this effort, the research undertaken by this project identified the limitations in existing technology and in the community's knowledge of underwater water acoustics that precluded achievement of the original objective. In the area of technology, the most obvious limitation was the limited speed of available computers that forced the use of cumbersome analog devices for signal processing and for beam steering. Techniques for controlling the location of the source dynamically and for the construction and for the operation of large-scale acoustic high-powered phased arrays were developed successfully. Unresolved issues that formed the basis for future research efforts related to signal coherence time, low-frequency acoustic cross sections of submarines, the development of reliable computation models for propagation loss, low-frequency acoustic absorption processes in ocean, and the development of algorithms that accounted for source Doppler. Project Artemis terminated in the mid-1960s.

1:25

2pUW2. The ancient and selective history of the development of high frequency, short range, high resolution active sonars. Chester M. McKinney (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

This is a brief history of the development of a broad class of active sonars which generally have most or all of these characteristics: relatively high operating frequency (35–1500 kHz), short range (a few meters to 1.0 km), high range resolution (1.0 m to a few cm), and high cross range resolution (a few degrees to 0.1°). Emphasis is on such sonars developed by several navies to find small objects such as naval seamines although there are other important civilian applications. Beginning with the experimental work of Paul Langevin (France) in 1916–18, the development of three generic types of sonars is traced: ahead searching, multibeam, electronic scanning sonars; continuous transmission frequency modulated sonars; and side-scanning sonars. Other types such as small boat sonars, sonars for manned and unmanned small submersibles, and 3D or camera type sonars are covered briefly. Basic research to support the development of all these sonars, primarily target strength and echo structure and bottom reverberation, is discussed only briefly as are other topics such as techniques for high cross-range resolution, target classification clues, and uses for wide bandwidth. Sonars that have proven to be historically significant are identified. The period covered is primarily 1916–1980.

1:50

2pUW3. Motivating the underwater acoustics community: 50s and 60s version. Fred N. Spiess (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD 0205, 9500 Gilman Dr., La Jolla, CA 92093-0205, fspiess@ucsd.edu)

In the decades immediately following the end of WWII our Navy was the only significant U.S. sponsor of underwater acoustics R&D. There were close, collegial, interactions among the academic and Navy laboratory researchers as well as between those groups and the operation-oriented Navy community. There was direct contact between the operating forces and the working levels in the

research organizations, exemplified by ComSubPac providing submarine services for research directly to the Navy Electronics Laboratory and the Marine Physical Laboratory. There was the ONR Deep Water Propagation Committee, where meetings started with a round table discussion of on-going research results before moving on to devising programs optimizing use of the research groups in the several academic and Navy laboratories. There was the Undersea Warfare R&D Planning Council in which senior officers from CNO met and shared their concerns with the directors of the Navy and academic laboratories, including critical review of the Navy's undersea warfare R&D budgets. This paper will explore a range of similar examples and show how they contributed to maintaining a high level of enthusiasm, collegiality and inventiveness resulting in major improvements in Navy sonar capabilities yesterday and providing perspectives for tomorrow.

2:15

2pUW4. Towed array history, 1917–2001. Stanley G. Lemon (Cheasapeake Science Corp., 1127B Benfield Blvd., Millersville, MD 21108)

Towed array research and developments over 84 years is examined from the inventions of Dr. Harvey C. Hayes during World War I to multiple arrays towed by oil exploration ships of today. A time line of the effort over the 80 years is presented with a discussion of the driving technical and military reasons for system evolutions. Military surface ship towed arrays from Hayes electric eel to the AN/SQR-14, SURTASS, ETAS, and APL/Johns Hopkins dual towed arrays are reviewed. Submarine array programs are described from the early 1960s through today. A prediction of future towed array developments is presented.

2:40

2pUW5. Oceanic ambient noise. William M. Carey (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Ambient Noise (AN) research during World War II was facilitated by the availability of calibrated instruments and motivated by the necessity to understand AN levels in coastal waters. The state-of-the-art was reviewed by Knudsen in 1948 and later by Pryce and Urick in 1954. After the war, AN research waned until the classic paper by Wenz (1962) marked its renaissance. By 1976, a vast literature existed on measurements, theory and computational methods, and in the 1980s AN was the second largest area of underwater acoustics. Initially, emphasis was on the omni-directional properties of the field and later on vertical/horizontal directionality and the shipping-component statistical characteristics. Measurement arrays characterized the statistics of the AN, its persistent directionality, and benchmarked attempts to predict the AN-levels. Interest in AN sources re-emerged around 1985 with the identification oceanic bubble ensembles as a major contributor to deep and shallow water noise fields. This overview discusses fundamental AN sources, measurements, and the physics of the air–sea interaction boundary layer. Bubble, spray, splash, and rain noise are examined and ranked in order of importance. The production, absorption, and transmission of sound by bubble clouds and layers produced by breaking waves are examined. [Work sponsored by ONR.]

3:05–3:20 Break

3:20

2pUW6. Advances in high-frequency underwater acoustics. Edward G. Liszka (Office of Naval Res., 800 N. Quincy St., Ballston Towers One, Ste. 907, Arlington, VA 22217-5660)

Significant progress has been made over the past 3 decades in understanding the acoustic phenomena that affect the operation of high-frequency (i.e., 10 kHz–60 kHz) sonars utilized by systems in the undersea environment such as weapons and autonomous underwater vehicles. The specific underwater acoustic research interests have somewhat paralleled and been motivated by the technology developments for these applications. In the late 1960s and early 1970s, experiments and model development focused on understanding the processes of signal propagation, scattering from the ocean medium (reverberation), and target reflection to predict signal levels for energy detection systems. Recently, systems have evolved to exploit more advancd signal processing concepts that have established a new set of scientific issues. Accordingly, underwater acoustics research has been focused on time/frequency/spatial distribution of energy, temporal, and spatial coherence of signals and background, and a broader scope of complex environments. The increased understanding has been captured in models representing the physical phenomena and rather detailed simulations of the underwater acoustic picture. This paper will review some of the key results in high-frequency underwater acoustics research that have contributed to the current state of knowledge and an indication of the direction for future research.

3:45

2pUW7. From matched filter to matched field tracking. Homer Bucker (SPAWARSYSCEN, 53560 Hull St., San Diego, CA 92152, bucker@nosc.mil)

An old example of use of a matched filter in underwater sound is the quadrature detector. The signal from hydrophone #j is split, and one channel is matched with a cosine wave and the other with a sine wave. These channels are assembled as the real and imaginary parts of the analytic signal F_j. The next application of a matched filter occurs in plane wave beam forming where the signals $\{F_j\}$ of an array are matched with signals $\{G_j(\varphi)\}$, which are expected if the incident signal is a plane wave with direction φ. The problem is expanded when a more precise form of $\{G_j\}$ is used. In general, $\{G\}$ will be a function of the location of the source. Therefore, in principle, the location of a sound source can be found by finding the coordinates of the best match between the F's and the G's (**matched field processing**). Several methods will be described showing how to do this in an efficient manner. Finally, there is further gain to be had by including many time values in matching the track of the acoustic source (**matched field tracking**). Examples will be shown illustrating results from sea tests.

4:10

2pUW8. Chemical sound absorption in the ocean. Frederick H. Fisher (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0701, fhf@mpl.ucsd.edu)

Two salts cause significant sound absorption in the ocean, small concentrations of magnesium sulfate ion pairs and boric acid. Below 100 kHz, MgSO4 produces sound absorption ~30 times greater than that of pure water. For boric acid below 1 kHz, the excess absorption in the Atlantic over pure water is ~300 times and in the Pacific ~150 times, the difference being lower pH at the sound axis in the Pacific than in the Atlantic (7.7 vs 8), the borate concentration being the same. For MgSO4, a four state dissociation model of Eigen and Tamm provides an explanation of sound absorption at 1 ATM from 10 kHz to 300 MHz. Remarkably, at 1000 ATM, this same model accounts for the observed large decrease in sound absorption (-66%) and the small increase in electrical conductance ($+10\%$). For boric acid, Simmons explained absorption in terms of a three state ion pair model involving a configurational change of the B(OH)4 ion from a tetrahedral to a planar trigonal structure B(OH)3 involving a large volume change of 36 cm/mole. The discovery, some history and the current status of these once ''anomalous'' absorptions will be discussed. [Work supported by ONR and NSF.]

4:35

2pUW9. Fifty years of measuring, modeling, and sometimes understanding the physics of sound in the sea. Helen M. Walkinshaw (Box 72, Peapack, NJ 07977, hmckayw@juno.com)

By the end of the 1940s naval experience in World War II and progress in undersea physics and oceanogoraphy had brought about recognition of low frequency as a means of passive detection. The discovery of the deep sound channel was published in 1948, defining its role in long range propagation. In January 1950, the National Research Council's Committee on Undersea Warfare recommended a study be made of these emerging concepts. What came to be known as the Hartwell Committee was formed. The Committee's report included specific recommendation of real-time narrow band analysis of radiated sounds and exploitation of low-frequency signals. In October, the Chief of Naval Research was directed ''to proceed with arrangements with Bell Laboratories to institute a program of research and development in the field of low frequency.'' This directive, shared with naval laboratories and affiliated universities, initiated an intensive program of sea experiments that continued over four decades. The measurements were paralleled by computer codes developed to analyze and deconstruct data, and eventually model propagation and arrival structure in realistic, range-varying ways. The sequence of the discovery and validation of features that define the nature of undersea sound as now understood are outlined and discussed.

5:00–6:00
Panel Discussion

Session 3aAA

Architectural Acoustics and Musical Acoustics: Integration of Synthesis Techniques with "Acoustic" Music

K. Anthony Hoover, Cochair
Cavanaugh Tocci Associates, 327F Boston Post Road, Sudbury, Massachusetts 01776

Richard H. Campbell, Cochair
Bang-Campbell Associates, Box 47, Woods Hole, Massachusetts 02543

Invited Papers

9:05

3aAA1. Good vibrations: Composed instruments integrating gestural control, and sonic display. Curtis Bahn (Rensselaer Polytechnic Inst., iEAR Studios, Dept. of the Arts, 110 8th St., Troy, NY 12180, crb@rpi.edu)

This paper examines electronic composed instruments which take as essential aspects of their design the full feedback loop of: acoustic sound production; physical interface/gestural sensing; computer interface (including sound synthesis/processing); and localized sonic display. Many issues relevant to architectural sonics are seen, including: the design of localized haptic/sonic feedback for the performer; the composition of relationships between gesture and sound production composing the sonic architecture of the body; and the inclusion of spherical speaker arrays to engage the natural acoustics of the hall. Conventional sound reinforcement techniques often draw acoustic instruments into an electronic sound space where artificial reverberation and other processing is added in an attempt to create a natural blend with electronic sound sources. This approach decorrelates instrumental gesture and original sound source, and most often conflicts with or undermines the acoustic design of the hall. The integration of sonic display into composed instruments reinforces the feedback loop between instrumental touch and sonic result, offers localized sonic feedback for the performers on stage, and creates a rich, acoustic ensemble blend. These extended instruments provide an extremely effective approach to integrating electronic musical performance into an acoustic context.

9:25

3aAA2. The "Concerto for Conductor:" A new musical form. Teresa Marrin Nakra (Artistic Director, Immersion Music, Inc., Boston, MA)

In May of 2001, at Boston's Symphony Hall, conductor Gil Rose walked out onto an empty stage with a coil of wire trailing from the back of his jacket. As he raised his hands in a starting gesture, the first chords of the Grieg Piano Concerto rang out. Subsequent gestures finished the initial flourish of the Grieg, and after several more bars of sampled orchestral and environmental music, orchestral musicians started slowly filtering onto the stage. The rest of the performance featured some very fun and energetic music making, reflecting both a sense of humor and a sense of inventiveness with a completely new form. This was the premiere performance of the world's first "Concerto for Conductor." The new musical form was the brainchild of this author, who wanted to develop a way to combine the "Conductor's Jacket" sensor interface with live musicians on-stage. Composer John Oswald, whose imaginative setting was highly praised by the Boston press, was chosen and commissioned because of his reputation for inventive use of sampled sound. Aspects of his composition, the electronic systems behind it, and the possibilities for future Conductor Concertos will be discussed in the public presentation of this paper.

9:45

3aAA3. New performance techniques using Max/MSP and Csound. Arran Lowe (Berklee College of Music, 150 Massachusetts Ave., Box # 1837, Boston, MA 02115)

New musical developments have often come about with the discovery or invention of new techniques. Software programming in the Max environment and the language of Csound can allow one to develop new techniques for real-time performance. Using the computer as a compositional tool, sound design device, signal processing source, and performer, one is able to accomplish musical goals that were not possible due to the limitations of human performance with previous tools. Combining live human performers and dynamically controllable software synthesis, it is possible to attain a synergy of very familiar and very new ideas. This lecture and performance demonstration will feature some of the new musical possibilities that have resulted from undergraduate research and a dissertation in software synthesis.

10:05

3aAA4. A versatile wavetable-based model for synthesis of piano tones. James W. Beauchamp, Hua Zheng, Bowon Lee (School of Music and Dept. of Elec. and Computer Eng., Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801, j-beauch@uiuc.edu), and Stephen Lakatos (Dept. of Psych., Washington State Univ., Vancouver, WA 98686)

A synthesis model has been developed based on short-time spectral analysis of a wide range of Yamaha Diskclavier notes performed at a variety of velocities and articulations. Frequency components are sorted by critical band, and each band is reduced to a wavetable controlled by the fundamental frequency and the rms amplitude envelope for that band. Inharmonicity is controlled by

appropriately stretching the fundamental frequency for each wavetable. In some cases, it is possible to combine critical bands into larger bands, thus improving the data reduction with only slightly reduced quality. In addition, transient noise, obtained by subtracting a full sinusoidal synthesis from the original signal, is used to improve attack and release authenticity. Preliminary results of listening tests indicate that on average subjects correctly identify full-critical-band-model synthetic tones only about 60% of the time when compared to original sounds. An intelligent computer synthesis instrument has been coded which allows production of arbitrary musical scores of high quality using parameter values dictated by note and velocity. However, the user may also vary certain parameters such as transient noise level, attack rate, and inharmonicity to achieve greater variety.

10:25–10:35 Break

10:35

3aAA5. The manipulation of audio dynamics toward both technical and artistic goals in recorded music. Alexander U. Case (Fermata Audio and Acoust., 117 Atlantic Ave., North Hampton, NH 03862, acase@fermataco.com)

The musical dynamics of recorded sounds must be made to fit within the audio dynamic range of the recording and playback equipment. This constraint on dynamic range is presented track by track in the multitrack recording studio and at every stage of audio signal processing up to and including the end user's system. Recorded orchestras must be made to sound good on clock radios. Audio compressors and limiters have been developed to help the recording engineer cope with this problem. Automatic, but careful, control of the signal gain during the course of music recording and playback is part of almost all music production styles, most notably pop music. In addition, these dynamics processors are used for purely aesthetic purposes. They shape the character, timbre, and emotional feel of a musical performance. These more innovative applications of compression become part of the music, creating effects that range from subtle to conspicuous. This paper reviews many compression techniques used in the production of recorded music and describes their technical and creative goals.

10:55

3aAA6. Synthesizer as processor: Flexibility in signal processing. Jan P. Moorhead (Berklee College of Music, 1140 Boylston St., Boston, MA 02215-3693, JMoorhead@Berklee.edu)

Creative use of the digital inputs available on many synthesizers offers a wide array of possibilities in digital signal processing. This provides the capabilities of standard rack-based processors as well as a great deal more. The most significant opportunities for new timbers come from modular synthesizers that allow you to design new instruments to suit your specific application. These techniques may easily be used in live performance as well as in the studio. When the unprocessed sound of the source is suppressed, it is possible to create entirely new timbers from standard acoustic or electroacoustic instruments. An advantage over typical MIDI-controlled synthesizers is that the performer need not necessarily learn a new interface or new techniques to optimize the expressiveness of the performance. Additionally, because the sound from the input instrument can be used as a controller for elements of the patch (in the synthesizer), the resulting output can be very responsive and expressive. In all, this provides a new class of nonstatic timbers for the performer, composer, engineer, and producer.

11:15

3aAA7. Using signal processing in real-time trumpet performance. Thomas J. Plsek and Daniel Brantigan (Berklee College of Music, 1140 Boylston St., Boston, MA 02215, tplsek@berklee.edu)

The real time use of signal processing for trumpet instrument performance provides an opportunity, as well as challenges, to create a type of hyper-instrument quite unlike most conceptions of the term. These include the expansion of the available sound palette, the problem of real-time parameter control, microphone techniques, and repertoire considerations. The various types of processing used includes reverb, pitch change, delays—short time and longer looping types, distortions, and combinations of the above. Normally, the processed signal is layered onto the acoustical output of the instrument, but by using such devices as the Yamaha Silent Brass one can diminish the presence of the acoustical signal of the instrument in the final output. By creatively managing the acoustic instrument and the electronic equipment, a vast array of musical resources become available to the performer enabling him/her to enhance existing performance environments, as well as find and develop new ones.

11:35

3aAA8. The integration of computer-generated sounds and acoustic sources during live electro-acoustic improvisation. Seth A. Cluett (30 E. Main St., Second Fl., Millers Falls, MA 01349) and Scott Smallwood (iEAR studios at Rensselaer)

A performance duo specializing in live electro-acoustic improvisation sought to create a balanced orchestration between live sound sources and computer generated sound. The sound sources were a Trinidadian steel-pan and four sheets of lead plate glass. The improvisation was limited to controlled, disciplined movements whose minimal nature exploited the natural acoustic properties of the sound sources. The sounds of the steel-pan and the glass were analyzed and algorithms were written that extracted acoustic information from the live sound. The analyzed sounds were then processed and re-introduced into the improvisation with dynamic control of the amplitude, parameters for, and amount of processing. Spatialization and integration of the acoustic and computer-generated sounds was considered and a multi-channel speaker system was used to introduce both into the performance space. This performance was undertaken for the 2001 International Computer Music Conference, in Havana, Cuba.

Session 3aAOa

Acoustical Oceanography and Archives and History: History of Acoustical Oceanography

Peter F. Worcester, Chair
Scripps Institution of Oceanography, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093

Chair's Introduction—7:55

Invited Paper

8:00

3aAOa1. History of acoustical oceanography. Robert C. Spindel (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, spindel@apl.washington.edu)

This lecture traces the history of acoustical oceanography from its beginnings during WWII (although it was not called Acoustical Oceanography at the time) to the present. We describe how during the Cold War the U.S. and Soviet Navy's efforts to improve the performance of their torpedoes and sonars revealed unknown ocean features and processes. We discuss the prominent role of the U.S. Navy's Office of Naval Research in bringing underwater acoustics and oceanography together, and fostering the science upon which Acoustical Oceanography is based. We show how the two fields eventually fused into a new discipline called Acoustical Oceanography, and how that combination has resulted in today's acoustic methods for observing the ocean on micro, regional, and global scales.

3a WED. AM

A NOTE ABOUT THE ASA HISTORY LECTURE SERIES

In 1997, the ASA Committee on Archives and History conceived a plan for a series of invited lectures on each of the technical areas of the Society which would memorialize the significant achievements and milestones of each of its technical committees during the first three quarters of the Society's first century.

With the cooporation of the technical committees, distinguished individuals are selected to review the history of their particular technical specialty and present a lecture which shows how that activity has developed and has contributed to the Society at large and to the broad field of acoustics as well.

The invited lecturers have been asked to prepare a written manuscript of their lectures which will be published in a commemorative book for the 75th Anniversary of the Society to be celebrated in 2004. The Archives and History Committee and the individual technical committees welcome comments and suggestions on both the History Lecture Series and on the proposed ASA Diamond Anniversary Book. Volunteers to assist the committees would be most welcome too. Contact Henry Bass, Chair, Committee on Archives and History, pabass@sunset.backbone.olemiss.edu

Session 3aAOb

Acoustical Oceanography: Acoustical Oceanography Prize Lecture

Peter F. Worcester, Chair
Scripps Institution of Oceanography, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093

Chair's Introduction—9:30

Invited Paper

9:45

3aAOb1. Surf zone bubble spectrometry: The role of the acoustic cross section. Timothy Leighton (Inst. of Sound and Vib. Res., Univ. of Southampton, Highfield, Southampton SO17 1BJ, UK)

Fifty years ago, there were doubts as to whether microbubbles existed in the sea at all. Herman Medwin settled the issue in the 1960s by showing that they were indeed present, and in very large numbers. He interpreted *in situ* measurements of backscatter and attenuation using the concept of the acoustic cross section. In essence that procedure forms the basis of the most popular ways of measuring microbubbles today, the two favored experimental layouts (measurement of attenuation and sound speed along a propagation path, and the acoustical resonator) being based on techniques pioneered by Medwin. In recent years attempts have been made to take these open ocean techniques and deploy them in the surf zone. Here the assumptions inherent in the standard form of the cross section can be violated. Specifically, the conditions may not relate to plane wave insonification of isolated bubbles undergoing steady state single-frequency linear pulsations in the free field in an incompressible liquid. This paper describes the attempts that have been made to produce cross sections which are not restricted by these limitations, and how these are exploited in determining the evolution, anatomy, and effects of bubble clouds. [The support of The Royal Society Leverhulme Trust is acknowledged.]

Session 3aEA

Engineering Acoustics, Physical Acoustics, Noise, Underwater Acoustics and Structural Acoustics and Vibration: Celebration of Miguel C. Junger's Contributions to Acoustics

Ira Dyer, Cochair
9 Cliff Street, Marblehead, Massachusetts 01945

David Feit, Cochair
Carderock Division, Naval Surface Warfare Center, 9500 MacArthur Boulevard, West Bethesda, Maryland 20817

Chair's Introduction—8:45

Invited Papers

8:50

3aEA1. Vibration damping layer performance enhancement with an intermittent bond. J. Garrelick (Cambridge Acoust. Assoc., Anteon Corp., 84 Sherman St., Cambridge, MA 02140, jgarrelick@anteon.com)

Having arrived in the U.S. in 1941 and upon graduating from the MIT in 1944, Miguel C. Junger's early contributions to structural-acoustics were predominantly structural. With war needs dominating, Miguel worked on aircraft engine testing at Wright Aeronautical Corp. and subsequently Naval ship shock mounts at Barry Controls. The clichéd ''can do'' attitude of these early war years may or may not have been the genesis of his much admired approach to technical problems of all stripes, a focus on fundamental physical principles and the utilization of rigor to extract the essence. In this talk, by example, I will attempt to illustrate how ''Miguel's acoustics'' has influenced and benefited his colleagues. The example is a problem recently confronted by the author, the effect of an intermittent bond on damping layer performance. Damping treatments are typically uniformly bonded, with any irregularity in converage deemed shoddy. However, this need not be the case and an intermittent bond may actually enhance performance. This will be demonstrated for an unconstrained layer with the intermittency represented by periodic gaps. Numerical results are compared for two solution approaches, one a somewhat plodding viscoelastic layer analysis, the other a rather highly idealized asymptotic representation, to capture the essence.

9:10

3aEA2. Airgun technology for undersea long-range acoustic detection systems. John V. Bouyoucos (Hydroacoustics, Inc., P.O. Box 23447, Rochester, NY 14692)

Miguel Junger and this author occupied a basement office at Harvard University's Acoustics Research Laboratory under the tutelage of Professor F. V. Hunt during the mid-1950's. Subsequently, Dr. Junger and his staff at Cambridge Acoustical Associates provided consulting support on numerous occasions to the evolution of hydroacoustic transduction technology carried out by the author and his engineering staff at General Dynamics Corporation and later at Hydroacoustics, Inc. in Rochester, NY. A key contribution was made by Dr. Junger and his associates to the evolution of the MULTIPULSE^{TM*} airgun array technology in defining the influence of the bubble cloud emitted from the vertical gun array on the azimuthal directivity of the transmitted broadband energy. This paper discusses the evolution of the vertical airgun transmit array and includes some far-field, at-sea measurements that generally support the analytical findings of Dr. Rudolph Martinez, one of Miguel Junger's colleagues at Cambridge Acoustical Associates, as referenced in paper 3aEA9 in this session, entitled ''Diffracting Open-Ended Duct of Finite Length: Flow Effects.''
*MULTIPULSE^{TM} is a registered trademark of Hydroacoustics, Inc.

9:30

3aEA3. Miguel C. Junger and the early Cambridge Acoustical Associates. Preston W. Smith, Jr. (61 Broad Reach, #M55B, Weymouth, MA 02191)

In 1955, Junger and this author (Smith) co-founded the partnership Cambridge Acoustical Associates (CAA), directly from our post-graduate studies at the Harvard Acoustical Research Laboratory. This association continued into 1959 when Smith left to join Bolt Beranek and Newman Inc. while Junger persevered with CAA. This paper presents an historical survey of Junger's professional activities in that early period of his life.

9:45

3aEA4. Miguel Junger, Cambridge Acoustical Associates, and the U.S. Navy. M. Strasberg (David Taylor Model Basin, NSWCCD, 9500 MacArthur Blvd., West Bethesda, MD 20817-5700)

Miguel Junger and his associates at Cambridge Acoustical Associates have had a long, active, and fruitful relationship with the U.S. Navy and Navy acousticians. In this talk we will review some of the papers and reports published in the open literature during the 1960s and 1970s, describing their work on subjects of interest to the Navy.

10:00

3aEA5. A practical side of M. C. Junger. Klaus Kleinschmidt (Cambridge Acoust. Assoc., Anteon Corp., 84 Sherman St., Cambridge, MA 02140)

This is a chapter in the life of a man whose mind is in part filled with a mist of symbols and equations but which has another part that can create very practical applications of theoretical acoustics. I am fortunate enough to have shared for many years in Miguel's prodigious intellectual output as well as his kindness and generosity. I will summarize the story of how he invented a product, basically made of dirt, about 50 years ago and of his participation in creating a unique and commercially successful acoustical material called SOUNDBLOX.

10:20–10:30 Break

10:30

3aEA6. A new model for an underwater explosion bubble. Thomas L. Geers and Kendall S. Hunter (Dept. of Mech. Eng., Univ. of Colorado, Boulder, CO 80309-0427, geers@spot.colorado.edu)

A new model for a moderately deep underwater explosion bubble is developed that integrates the shockwave and oscillation phases of the motion. A hyperacoustic relationship is formulated that relates bubble volume-acceleration to far-field pressure profile during the shockwave phase, thereby providing initial conditions for the subsequent oscillation phase. For the latter, new equations are derived that include wave effects in both the external liquid and the internal gas. The equations are then specialized to the case of a spherical bubble, and bubble–surface displacement histories are calculated for dilational and translational motion. Agreement between these histories and experimental data is found to be good. [Research sponsored by the DTRA and ONR.]

10:55

3aEA7. Miguel C. Junger, an inspirational mentor in structural acoustics. David Feit (Carderock Div. Naval Surface Warfare Ctr., 9500 MacArthur Blvd., West Bethesda, MD 20817)

When Lord Rayleigh wrote his book, *The Theory of Sound* in the 19th century, he treated vibrations of structures and sound radiation in two separate volumes. In the last half of the 20th century, the field of structural acoustics, i.e., the theory of acoustic radiation and scattering from elastic structures, was developed in response to Navy needs. Miguel C. Junger is one of the seminal contributors to this field of structural acoustics, and has been the inspirational mentor to myself and others associated with him at Cambridge Acoustical Associates. This presentation discusses how this mentoring proceeded, the resulting research derived from this, and the circumstances leading to our book *Sound, Structures and their Interaction* originally published by MIT Press, and now published by the Acoustical Society of America.

11:15

3aEA8. Sound and structural destruction. James E. Barger (BBN Technologies, 10 Moulton St., Cambridge, MA 02138)

Our honoree is well known for his, and David Feit's, book on the interaction of sound with structures. In this work, equations representing both sound and structure are linearized. Results apply to most cases, even when sound pressures exceed safe hearing levels. But there are occasions when it is wanted to destroy the functionality of a structure by a directed shock wave in its surrounding fluid. This type of interaction exceeds yield stress in the structure, causing permanent deformation. This paper describes both an explosive and a nonexplosive underwater shock generation device and shock propagation to a 0.8 m diam metal sphere. The damage is computed by finite element methods, and compared to actual damage. Insights are found from the linear analysis.

Contributed Papers

11:35

3aEA9. Diffracting open-ended duct of finite length: flow effects. R. Martinez (Cambridge Acoust. Assoc., Anteon Corp., 84 Sherman St., 3rd Ste., Cambridge, MA 02140, rmartinez@caa.atinc.com)

Dr. Junger once addressed the problem of diffraction by ducts of finite length using a variational approach that supplied an approximate solution (personal communication, ca. 1986). His medium was quiescent. The speaker later tackled the problem numerically and more rigously [J. Acoust. Soc. Am. **87**, 523–531 (1990)], and provided an *a posteriori* check on that new solution via number of power-related quantities. The current presentation extends our earlier analysis to include the effects of a freestream on the acoustics of a free-flooded unbaffled duct, which in this context becomes a ring wing insonified from within. The theory's physical content includes the ring wing's shed and trailed vorticity systems, which are formally coupled to the acoustic field over the duct's surface. The presence of vorticity makes the standard boundary-element approach numerically unstable for reduced frequencies of $O(10)$ (reduced frequency being Strouhal number based on the ring wing's half-length). Our solution technique restores numerical stability through a reformulation of unsteady lifting-surface theory to be a Fredholm integral equal of the second kind. The presentation concludes with general remarks on diffraction theory for thin two-sided objects, particularly on the use of reciprocity in the calculation of diffracted nearfields for soft screens.

11:50

3aEA10. Scattering by tilted cylindrical shells in water above the coincidence frequency. Philip L. Marston, Florian J. Blonigen (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814), and Scot F. Morse (Western Oregon Univ., Monmouth, OR 97361)

Experimental, theoretical, and computational results indicate that flexural waves on bluntly truncated tilted cylindrical steel shells are a major contributor to the backscattering of sound at high frequencies over a wide range of tilt angles. Important contributions are associated with meridional rays and with helical rays [S. F. Morse *et al.*, J. Acoust. Soc. Am. **103**, 785–794 (1998); S. F. Morse and P. L. Marston, J. Acoust. Soc. Am. **106**, 2595–2600 (1999); S. F. Morse and P. L. Marston, IEEE J. Ocean. Eng. **26**, 152–155 (2001)]. Successful development of a ray model for helical ray scattering contributions required the computation (at high frequencies) of the anisotropy of the flexural-wave velocity and radiation-damping parameters [F. J. Blonigen and P. L. Marston, J. Acoust. Soc. Am. (accepted for publication)]. In backscattering experiments with short and long tone bursts, the associated meridional and helical ray contributions are typically much larger than the scattering estimated for diffraction by a rigid tilted cylinder of the same size. The importance of the elastic response of shells was recognized in the early work of Junger [M. C. Junger, J. Acoust. Soc. Am. **24**, 366–373 (1952)]. [Work supported by ONR.]

WEDNESDAY MORNING, 5 DECEMBER 2001 ROOMS 203/204, 9:00 TO 11:30 A.M.

Session 3aED

Education in Acoustics: Low Cost Laboratory Experiments

Daniel R. Raichel, Chair
2727 Moore Lane, Fort Collins, Colorado 80526

Chair's Introduction—9:00

Invited Papers

9:05

3aED1. Low cost laboratory experiments in musical acoustics. Thomas D. Rossing (Northern Illinois Univ., DeKalb, IL 60115)

For a number of years we have offered a one-credit Acoustics Laboratory course, which is available to students taking our introductory ''Acoustics, Music, and Hearing'' course as well as to other students. We have developed over 500 acoustics experiments, many of which can be done with low-cost materials or equipment, that are described in our manual *Acoustics Laboratory Experiments*. We offer the students the option of taking the materials home (''lab in a bag'') to do the experiments, where they will often attract the attention of parents or roommates. We describe some of the experiments that can be done with low-cost materials or equipment.

9:25

3aED2. Inexpensive laboratory experiments for nonscience students. Uwe J. Hansen (Dept. of Phys., Indiana State Univ., Terre Haute, IN 47809)

Students who experience acoustics rather than merely being exposed to concepts universally gain insights not otherwise available. The experience of doing rather than simply hearing greatly enhances the excitement associated with learning. With that understanding, it seems almost imperative to include a laboratory component in every acoustics course. Frequently cost precludes the use of sophisticated equipment in multiple sets of identical experiments, however, sequential use of different laboratory setups provides a simple solution to that problem. Furthermore, there are many simple devices which lend themselves to quantitative measurements, and there is available relatively low-cost computer software for rather sophisticated experiments. Among experiments to be discussed and illustrated are graphic wave addition, complex wave synthesis using a polyphonic keyboard and a computer based digital oscilloscope, standing transverse waves on a spring, standing longitudinal waves in an air column, spectral analysis using inexpensive computer software with a sound card in a PC, normal mode mapping with Chladni patterns and phase shift observation in Lissajou figures, and sound pressure level measurements.

9:45

3aED3. Acoustics and noise control curriculum development at Wayne State University. Sean Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

This paper presents an overview of the curriculum development on acoustics, vibration, and noise control at Wayne State University. Special attention is given to the applications of acoustics and noise control principles to NVH (noise, vibration, and harshness) problems that are vital to the automotive industry. This emphasis is important in the curriculum development, given the fact that Wayne State is an urban university and many students are practicing engineers working at the local industries. Many hands-on experience and laboratories are included in the courses. Computer projects on the simulation of sound radiation from finite objects and on the designs of enclosures with specified transmission losses are incorporated in the course development. Research topics carried out at the Acoustics, Vibration, and Noise Control Laboratory are also integrated into the curriculum.

10:05

3aED4. The evolution of test instrumentation: From rack and stack to plug and play. Martin Alexander (Bruel & Kjaer, 63 Passaic Ave., Summit, NJ 07901, marty.alexander@bksv.com)

Electronic test instrumentation for acoustics and vibration measurement has seen radical changes during the final decade of the 20th century. From analog to digital to PC based, the performance, flexibility and ease of use have risen while the price per performance has declined. Taking advantage of the rapid developments in PC operating systems, mass marketed microprocessors, and the professional audio industry, todays instrumentation frequently has more in common with both the recording studio and the home computer than with the physics laboratory. The future promises even more as this evolution continues. Two examples, vehicle vibration analysis and electroacoustic testing, are used as examples to follow the development of test platforms. The advantages (and disadvantages) of software-based testing are discussed, and some insights of what the new decade might offer are provided.

10:25

3aED5. Optimizing selection of acoustical measurement equipment to meet educational needs. Noland Lewis (ACO Pacific, Inc., 2604 Read Ave., Belmont, CA 94002, acopac@acopacific.com) and Daniel R. Raichel (CUNY Grad. Ctr., Fort Collins, CO 80526)

As part of their learning experience, students of acoustics need to know and understand measurement techniques, particularly those involving sound pressure levels, sound intensities, spectrum analyses, and signal generation. It is important to realize that students not only should obtain hands-on experience, but also gain a keen appreciation for significant factors affecting accuracy and long-term stability of the equipment being chosen for use. The requisite equipment may not be inexpensive but some economies can be effected by judicious selection of devices without seriously compromising the validity of experimental results. For example, Type 2 microphones, rather than Type 1, may suffice to yield revealing trends in measurements of ''real life'' situations; and due consideration must be given to the ruggedness of equipment with respect to storage conditions and environmental exposure during use. Fundamental experiments include measurements of SPL, spectrum analyses, measurement of room reverberation time, testing of loudspeaker response including THD, sound absorption effects, etc. Selection of equipment with due regard to costs can be eased by the purchase of a ''university package'' of sensing devices that can be applied with FFT and sound generation cards and special purpose computer programs.

Contributed Papers

10:45

3aED6. Computer simulation and virtual experiments for architectural acoustics education. Ralph T. Muehleisen (Dept. of Civil, Environ., and Architectural Eng., Univ. of Colorado, Boulder, CO 80309)

Computer simulation and virtual computer experiments are one way of introducing students to concepts without the high cost of traditional experimental equipment. Computer simulation can show students the effects of room reverberation on different types of acoustic signals including speech and music and to show the noise control effects of absorptive materials. Computer based digital filters can be used with audio files to auralize the acoustic transmission loss of different materials and constructions. In this presentation the use of computers for simulation and virtual experiments in architectural acoustics is discussed. A variety of modern computer programs for acoustic prediction and simulation of room acoustics and noise control are reviewed.

11:00

3aED7. Determining the temperature dependence of Young's modulus using a tuning fork. Earl D. Blodgett (Dept. of Phys., Univ. of Wisconsin, River Falls, River Falls, WI 54022)

We describe a simple experiment that illustrates the use of acoustic resonance to determine the temperature dependence of an elastic constant. Instead of a traditional crystal that has been carefully cut and polished to provide a high Q, the specimen is an ordinary tuning fork. The resonant frequency of a tuning fork changes with temperature because of changes in Young's modulus [D. W. Schindel *et al.*, J. Acoust. Soc. Am. **102**, 1296–1308 (1997)]. In the stainless steel fork used in a previous undergraduate research study [L. Bates *et al.*, J. Undergrad. Res. Phys. **18**, 9–13 (1999)], the resonant frequency decreased by 0.018 54 Hz for each 1 °C increase in temperature. That experiment was carried out over a range of temperature from 25 °C to 80 °C using a simple oven. As the frequency change is not very dramatic, it is very important to have either a method of very precisely measuring the resonant frequency or one must extend the temperature range to get a sufficiently large total shift in frequency. We describe several options for tracking the resonant frequency.

11:15

3aED8. Star in a jar. Roman G. Eskue (Willcox High School, 240 N. Bisbee Ave., Willcox, AZ 85643)

The hypothesis that sonoluminescence can be produced using a signal generator, an acrylic chamber, piezoelectric speakers, and water was tested. Proving this hypothesis took many experiments, most of which failed. For the first five experiments the speakers were connected to the outside of the acrylic chamber and wineglass. The chamber and glass were filled with different types of water: tap water, distilled water, chilled distilled water, and a chilled mixture of 80% distilled water and 20% glycerin. When the ultrasonic sounds were produced and bubbles were introduced in the water, nothing happened. In the sixth experiment the speakers did not fit well enough in the wineglass to produce enough resonating sound in the water. Finally, the speakers were placed in the acrylic chamber with the 80/20 mixture. As the signal generator was adjusted to a frequency of 28.3 KHz, there it was, that tiny blue light, the Star in a Jar. The hypothesis was correct: sonoluminescence can be produced using a signal generator, an acrylic chamber, piezoelectric speakers, and water.

WEDNESDAY MORNING, 5 DECEMBER 2001 ROOMS 118/119, 8:00 TO 11:20 A.M.

Session 3aNS

Noise: Ray-Tracing Software for Predicting Community Noise from Industrial Facilities I

Frank H. Brittain, Chair
3011 Triumph Drive, Alameda, California 94501

Chair's Introduction—8:00

Invited Papers

8:05

3aNS1. Overview of predicting community noise from industrial facilities using ray-tracing software. Frank H. Brittain (Bechtel Corp., 50 Beale St., San Francisco, CA 94105, fhbritta@bechtel.com)

Predicting community noise from industrial facilities is vital to both assessing the environmental impact and designing plants to meet community noise limits. Predicting noise levels with sufficient accuracy requires a thorough understanding of not only the software package used, but also the whole process. Just knowing how to input data to the software is never sufficient. Atmospheric effects, noise sources, the modeling process, and the assumptions built into the software must also be clearly understood. Until recently, community noise predictions have most commonly been made using a spreadsheet. This is convenient and simple, but it is difficult to adequately include many effects, such as screening and reflection of noise by both barriers and other equipment. Commercially available ray-tracing software allows more precise predictions of community noise, including screening and reflections. However, the accuracy of predictions depends primarily on the quality of the input to the software. As a start to the process of developing the understanding needed, this paper reviews the basic approach used for all community noise predictions, alternative prediction methodologies, capabilities of typical software packages, and the technology typically built into ray-tracing software. Developing models, and some of the problems often encountered are also identified and reviewed.

8:35

3aNS2. Industrial noise and noise maps for cities—An application of ray tracing techniques. Wolfgang Probst (Graefelfinger Str. 133A, D-81375 Munich, Germany, wolfgang.probst@datakustik.de)

It is shown how extended sources like radiating openings or surfaces, large area sources or extended line sources can be modeled for prediction purposes even if pure geometric models based on ray tracing are used. The limit of ray tracing technology comes into play where wavelengths and the dimension of reflecting or radiating objects are similar. Examples are presented using conical beam method and projection method. Diffraction is described by a barrier attenuation of the direct ray that is frequency dependent and that is based on the well known and experimentally adapted Maekawa formulas. If many arbitrary oriented diffracting and reflecting objects are crossed by the ray it may be necessary to add the contribution of more possible paths. Ray tracing models have their limitations like all other models used for noise prediction. On the other side it is not feasible to solve the wave equation using the environment as it exists in large and complex models. With this contribution some of these limitations are outlined and it is shown that these models can be used effectively for the prediction of noise levels in the vicinity of industrial sites.

9:05

3aNS3. Mechanisms affecting the propagation of sound through the atmosphere. Michael R. Stinson, Gilles A. Daigle, and David I. Havelock (Inst. for Microstructural Sci., Natl. Res. Council, Ottawa, ON K1A 0R6, Canada)

This talk will survey the various factors that affect the propagation of sound through the atmosphere. Ray-tracing software intended to predict community noise must accommodate these factors in some approximate yet realistic manner. The atmosphere is neither static nor uniform. Scattering by turbulence and refraction due to temperature and wind speed gradients must be considered. The impedance of the ground surface, the local topography (hills and valleys), and the presence of obstacles such as buildings have an important effect on the sound field. Barriers can reduce noise levels significantly, but the presence of turbulence and gradients can diminish their effectiveness.

9:35

3aNS4. Quirks, problems, and work-arounds with ray-tracing software. Marlund Hale (Adv. Eng. Acoust., 663 Bristol Ave., Simi Valley, CA 93065, noisedoc@aol.com)

More and more, acoustical consultants and those in industry are turning to commercially available community noise prediction software to assist in assessing and mitigating community and environmental noise impacts. End users come from varied backgrounds that range from the novice to the expert in noise prediction or using noise ray-tracing software. What could start out as an expectation of a faster and more effective method of analyses, often turns into disappointment and frustration, if not total exasperation. Furthermore, hours upon hours of wasted effort and budget overruns can occur while learning to effectively use the software package chosen. This typically stems from a lack of adequate tutorials or training, apparent errors in the software, poor documentation, undocumented restrictions, lack of ease in learning and using the software, etc. This paper identifies some of the problems often encountered while modeling industrial facilities, propagation path influences, and receivers. In addition, some problems found in basic acoustic modeling are also identified and reviewed. In certain instances, work-arounds have been found, which allow the modeling process to continue with reasonable accuracy.

10:05–10:20 Break

10:20

3aNS5. Elements of accuracy/inaccuracy in outdoor sound level predictions using ray-tracing models. Robert Putnam (Siemens Westinghouse, 4400 Alafaya Trail, Orlando, FL 32826)

Sound emissions play a vital and often pivotal role in a broad range of issues in the design of modern power generating facilities. The modeling of those sound emissions must be as accurate as possible in order to minimize the cost of noise control measures. This paper presents an overview of the principal components of the expected accuracy in modeling the sound emissions of sources to far field outdoor positions, including source data bounds. The relative magnitudes of the various elements of this accuracy will be discussed, and an assessment of practical expectations of achievable accuracy will be included. A comparison will be presented of preliminary results of a round robin survey involving several different computer models and a proposal will be offered for additional round robin testing.

10:50

3aNS6. Community noise modeling of industrial facilities. David Parzych (Power Acoust., Inc., 12472 Lake Underhill Rd., #302, Orlando, FL 32828)

Community noise prediction is often based on point source radiation models. Sound radiation models are exercised to account for attenuation due to atmospheric absorption, ground effects, barrier effects, and sound reflections. Often, industrial facility noise models ignore obvious factors, such as partial screening or partial reflections of large sound sources, directivity effects, and physical source dimensions as well. This paper addresses sound modeling issues including input data resources, dealing with large radiating surfaces, directivity, and decisions on whether to use simple screening models or sophisticated commercial software. Large equipment modeled as a point source is discussed and how point source noise models can be adapted to accurately estimate sound from large sources.

A panel discussion will be held in Session 3pNS

Session 3aPAa

Physical Acoustics: Scattering and Propagation

Kerry W. Commander, Chair
Coastal Systems Station, Code R21, Panama City, Florida 32407

Contributed Papers

8:00

3aPAa1. Nobelists in acoustics. Robert T. Beyer (Dept. of Phys., Brown Univ., Providence, RI 02912, beyer@physics.brown.edu)

The only Nobel Prize awarded specifically in acoustics was that to Georg von Bekesy in 1962 in medicine and physiology for his work on the ear, but there are at least ten other winners who did significant work in acoustics. The most famous and prolific of these was, of course, Lord Rayleigh, but the others include Pierre Curie (piezoelectricity), Peter Debye (the ultrasonic beam as a diffraction grating), C. V. Raman (mathematical analysis of the Debye–Sears effect), Albert Einstein (velocity dispersion of sound in gaseous mixtures), and Lev Landau (sound absorption in liquid helium). Their acoustical work and that of a number of other Nobel laureates will be discussed.

8:15

3aPAa2. Measured and predicted acoustic attenuation in binary gas mixtures. Sally G. Ejakov (Ctr. for Sensor Material, Michigan State Univ., 303 P/A Bldg., East Lansing, MI 48824-1116), Scott Phillips (Commercial Electron., Broken Arrow, OK 74012-2838), Yefim Dain, Richard M. Lueptow (Northwestern Univ., Evanston, IL 60208), and Jacobus H. Visser (Ford Res. Lab., Dearborn, MI 48121-2053)

Acoustic attenuation in a gas mixture results from the combined effects of molecular relaxation and the classical mechanisms of viscosity and heat conduction. Acoustic attenuation was measured in methane–nitrogen mixtures for a wide range of concentrations in a large vessel using a pulsed technique. The signals from four pairs of transducers, each operating at a different frequency, were measured as a function of the distance between the transducers. The pressure in the vessel was varied to cover a wide range of f/p. A correction for diffraction for each transducer pair was applied to the data resulting in overlap between the data from different transducers at the same f/p. The pure methane measurements are in good agreement with published results. Measurements for mixtures of methane and nitrogen agree well with calculations that include both relaxational and classical attenuation [J. Acoust. Soc. Am. **109**, 1955 (2001)]. Similar measurements were completed for mixtures of carbon dioxide and nitrogen. Calculations to predict the attenuation for these mixtures are ongoing. [This work was funded in part by the State of Oklahoma.]

8:30

3aPAa3. Angular Doppler effect in the scattering of a helicoidal wave by a rotating target. Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

The phase of an acoustic helicoidal wave advances linearly with angle around a propagation axis at a rate characterized by an angular index m. For $m=1$ and 2, the phase advances are 360 deg and 720 deg, respectively, for an angular advance of 360 deg. There is null in the amplitude along the beam axis. A transducer for approximating such a helicoidal beam near the axis has been demonstrated [B. T. Hefner and P. L. Marston, J. Acoust. Soc. Am. **106**, 3313–3316 (1999)]. In the new work, the scattering by small objects that rotate around the helicoidal beam axis is considered. The wave scattered to a receiver on the axis will be shifted in frequency. The shift is constant and is given in magnitude by the product of the target's rotation rate and the helicoidal wave's angular index m. The shift is a consequence of the angular dependence of the phase of the helicoidal wave and the sign of the shift depends on the sign of m and the direction of the rotation of the target. A similar angular Doppler effect should also be present in the scattering of electromagnetic helicoidal waves. [Work supported by ONR.]

8:45

3aPAa4. Backscattering of acoustic waves in a cavity with random variations in area and sound speed. Mark J. Beran and Shimshon Frankenthal (Faculty of Eng., Tel Aviv Univ., Ramat Aviv, Israel, beran@eng.tau.ac.il)

The transmission of an acoustic signal in a cavity with random variations in area and sound speed is discussed. We first review the case of a single propagating mode and study the statistical properties of the forward and backward propagating waves in a cavity with only random sound-speed fluctuations. It is then shown that adding the effect of random variations in cavity area does not change the form of the basic statistical equations and previously obtained solutions may be used. The only change is in the characteristic length scale of the solution. We next consider the case of two propagating modes when the intermodal coupling is small. It is shown how the statistical hierarchy obtained in the single mode case may be used to obtain the effects of mode interaction.

9:00

3aPAa5. Reradiation of acoustic waves from circumferential waves on elastic shells. A. Claude Ahyi, Hui Cao, P. K. Raju (Dept. of Mech. Eng., Auburn Univ., Auburn, AL 36849-5341), and H. Überall (Phys. Dept., Catholic Univ. of America, Washington, DC 20064)

We consider evacuated thin semi-infinite shells immersed in a fluid, either of cylindrical shape with a hemispherical endcap, or formed two-dimensionally by semi-infinite parallel plates connected by a semi-cylinder. The connected shell portions are joined in a steady fashion although with discontinuous curvature. Acoustic waves are considered axially incident onto the curved portion of the shell, where they, at the critical angle of coincidence, generate Lamb and Stoneley-type waves in the shell. A computer program was written, based on a previous investigation by Cao *et al.* [Chin. J. Acoust. Soc. Am. **14**, 317 (1995)] in order to computationally visualize the waves in the fluid that have been reradiated by the circumferentially propagating shell waves at the critical angle. The results will be used in order to interpret experimental results in which the reradiated waves are optically visualized by the Schardin–Cranz Schlieren method; the latter experimental results will be reported elsewhere.

9:15

3aPAa6. A higher-order on-surface radiation condition derived from an analytic representation of a Dirichlet-to-Neumann map. David C. Calvo, Michael D. Collins, and Dalcio K. Dacol (Naval Res. Lab., Washington, DC 20375)

An approximate solution to a scattering problem can be obtained efficiently using an on-surface radiation condition (OSRC) [G. A. Kriegsmann *et al.*, IEEE Trans. Antennas. Propag. **35**, 153–161 (1987)], which is an approximation of the Dirichlet-to-Neumann map. A higher-order

OSRC is derived using an analytic representation of the Dirichlet-to-Neumann map for a circle, which involves a Hankel function with a tangential operator appearing in the index. This operator formalism is related to the well-known square root of an operator that arises in the parabolic equation method and is asymptotic to the square root in the high-frequency limit. The Hankel function is approximated with a rational function to obtain an OSRC for the circle. An OSRC for a more general convex object is obtained by fitting osculating circles to the surface and allowing the coefficients of the rational approximation to depend on curvature. Since the computation time increases only linearly with frequency, accurate solutions are obtained with substantial computational savings in comparison with other approaches. The OSRC can be constructed to any desired accuracy in scattering angle and surface curvature. [Work supported by ONR.]

WEDNESDAY MORNING, 5 DECEMBER 2001 ROOM 222, 9:45 TO 11:45 A.M.

Session 3aPAb

Physical Acoustics: Material Properties

Ronald E. Kumon, Chair
National Institute of Standards and Technology, 325 Broadway, MS 851, Boulder, Colorado 80305-3328

Contributed Papers

9:45

3aPAb1. On the use of finite element method in resonant ultrasound spectroscopy. Hang Guo and Amit Lal (SonicMEMS Lab., Dept. of Elec. and Computational Eng., Univ. of Wisconsin–Madison, 1415 Engineering Dr., Madison, WI 53706)

The Rayleigh–Ritz method is conventionally used in Resonant Ultrasound Spectroscopy (RUS) for the determination of the elastic constants of materials. It limits the application of RUS only for samples in simple shapes. In order to use RUS to characterize the complicated MEMS structures, the Finite Element Method (FEM) in RUS has been studied in this paper. Starting from Hamiltonian's principle, FEM is compared with the Rayleigh–Ritz method for RUS application. As numerical tests, the accuracy of FEM is checked by using an isotropic quartz cube and an anisotropic corundum rectangular parallelepiped. The results show that for isotropic samples FEM converges very fast and resolution of calculated natural frequencies can be as high as 0.4%. Then FEM is used to calculate natural frequencies and make mode analysis of a micromachined silicon nitride membrane achored on silicon substrate. Experimentally, the natural frequencies for those coupled membrane modes identified by FEM have been measured by RUS and used to determine Young's modulus of silicon nitride, which is 197 GPa and verified by comparing with measured results from the Load-Deflection method. This illustrates that RUS combined with FEM can be used as a new tool for Non-Destructive Testing of thin films in MEMS structures.

10:00

3aPAb2. The complex phase gradient method: A new tool for the investigation of immersed elastic plate angular resonances. Olivier Lenoir, Jean-Marc Conoir, and Jean-Louis Izbicki (LAUE, UMR CNRS 6068, Universite du Havre, Pl. R. Schuman, 76610 Le Havre, France)

The Phase Gradient Method applied to the complete characterization of the angular resonances of an immersed elastic plate, i.e., the angular poles of the plate reflection coefficient, was proven to be efficient, as far as the pole imaginary parts are not too large when compared to their real parts. This method consists of plotting the reflection coefficient phase partial derivative with respect to the sine of the incidence angle, considered as real, vs incidence angle. At the vicinity of an angular resonance, the plot exhibits a Breit–Wigner shape, whose minimum is located at the pole real part and whose amplitude is the inverse of its imaginary part. However, when the imaginary part is too large, this method does not give its value with a sufficient accuracy, compared to the exact calculation of the root in the complex angle plane. An improvement consists in plotting in 3D, in the complex angle plane, the angular phase derivative with respect to the real part of the sine of the incidence angle, considered as complex. When an angular pole is reached, the 3D curve shows a clear cut transition of very large amplitude, whose position is easily obtained.

10:15

3aPAb3. Introduction of frequency, angular, longitudinal, and transverse quality factors to characterize the resonances of a water-loaded elastic plate. Olivier Lenoir, Jean-Marc Conoir, and Jean-Louis Izbicki (LAUE, UMR CNRS 6068, Universite du Havre, Pl. R. Schuman, 76610 Le Havre, France)

The frequency derivative of the phase of the immersed elastic plate reflection coefficient is related to its derivatives with respect to the longitudinal and transverse phase velocities in the plate and to the phase velocity in the surrounding fluid. In this relation, products, composed of one of these variables and of the corresponding phase partial derivative, appear. At a resonance frequency, the frequency phase derivative amplitude corresponds to the inverse of the resonance half-width; the frequency-frequency phase derivative product can be identified to a frequency quality factor, denoted as Qx. By analogy, all the other products correspond to quality factors. It is shown that the product assigned to the fluid phase velocity corresponds to an angular quality factor, denoted as Qy. The products assigned to the longitudinal and tranverse phase velocities are naturally qualified as longitudinal and transverse quality factors, denoted as Ql and Qt. It is established that the sum of Qx and Qy equates the sum of Ql and Qt. Moreover, each of these quality factors is shown to be related to a stored kinetic energy, associated either to a normal displacement, to a parallel displacement, to a longitudinal or to transverse displacement.

10:30

3aPAb4. Sound absorption in titanium at 25 kHz. F. H. Fisher and C. C. Hsu (Marine Physical Lab., Scripps Inst. of Oceanogr., Univ. of California, San Diego, CA 92093-0701)

In the process of studying sound absorption in magnesium sulfate solutions as a function of pressure, a 100 L titanium sphere (a spare ballast tank for the WHOI researc submarine ALVIN) was used up to 300 atm. In calibrating the losses due to the sphere with pure water, Hsu found a resonant mode at 25 kHz (using a HP microHz synthesizer) with a very

long decay at 1 atm and 25 °C; a straight line 50 dB decay of 70 s. This means that ultrasonic absorption in the titanium sphere (4 Al, 6 Va alloy) filled with water is 5.5 exp(−7) N/cm. This is four times greater than that of water at 25 kHz. Therefore, subtracting the absorption of water yields an upper limit for absorption in titanium of 4.1 exp(−7) N/cm. The 0.9 m sphere was suspended from two loops of piano wire in a steel vacuum chamber. The resonance ($Q = 955\,000$) required considerable attention to insuring pressure, temperature, and frequency control, as well as the usual degassing of the liquid. Discussion of results and details will be presented. [Research supported by ONR.]

10:45

3aPAb5. Periodicity of linear and nonlinear surface acoustic wave parameters in the (111) plane of cubic crystals. A. Lomonosov, P. Hess (Inst. of Physical Chemistry, Univ. of Heidelberg, Im Neuenheimer Feld 253, 69120 Heidelberg, Germany), R. E. Kumon (Natl. Inst. of Standards and Technol., Boulder, CO 80305-3328), and M. F. Hamilton (Univ. of Texas, Austin, TX 78712-1063)

Previous numerical and experimental studies of surface acoustic waves (SAWs) in the (111) plane of cubic crystals have shown that linear properties, including the wave speed and the direction of power flow, exhibit sixfold symmetry. However, recent measurements by Lomonosov and Hess of finite-amplitude SAW pulses propagating in opposite directions in the (111) plane of crystalline silicon have demonstrated that the same periodicity does not hold for the nonlinear distortion. Calculations based on a model for nonlinear SAWs in anisotropic media [Hamilton *et al.*, J. Acoust. Soc. Am. **105**, 639–651 (1999)] indicate that complex-valued nonlinearity matrix elements, which describe the coupling between harmonics, have sixfold symmetry in magnitude but only threefold symmetry in phase. As a result, pulses traveling in opposite directions exhibit different nonlinear phase shifting between harmonics that give rise to the distinctly different types of observed distortion. Threefold symmetry of nonlinearity is also predicted for other cubic crystals besides silicon. Additional computations show that the complex-valued eigenvalues and eigenvectors of the linearized equations (physically corresponding to the depth decay coefficients and component amplitudes, respectively) also have sixfold symmetry in magnitude but only threefold symmetry in phase. [Discussions with A. P. Mayer are gratefully acknowledged.]

11:00

3aPAb6. Multichannel resonant scattering theory applied to the determination of polarization state of surface waves. Emilian Danila, Pascal Rembert, and Jean-Louis Izbicki (LAUE UMR CNRS 6068, Universite du Havre, Pl. R. Schuman, 76610 Le Havre, France)

In the case of scattering by a fluid filled cavity in elastic solid, the scattering appears into two channels, the P-channel and the S-channel. Then the scattering S-matrix is a block diagonal matrix. The elements of this matrix characterize the scattering with or without change of polarization state. From the multichannel resonant scattering theory it results that resonance widths are the sum of two partial resonance widths. This is an expression of the energy conservation law in the scattering process. It is difficult to analyze the distribution of the energy over the P respectively S channel, in the complex plane associated to the poles of S-matrix elements, especially in the neighborhood of the repulsion points on the dispersion curves. A complete analysis is then made using the complex planes associated to the zeros of the numerators of the S-matrix elements.

11:15

3aPAb7. Propagation of acoustic waves through finite solid superlattices. Betsabe Manzanares-Martinez (Posgrado en Ciencias (Fisica), Universidad de Sonora, Calle Rosales y Blvd. Luis Encinas, Hermosillo, Mexico) and Felipe Ramos-Mendieta (Centro de Investigacion en Fisica de la Universidad de Sonora, Hermosillo, Mexico)

The transmission of longitudinal waves through solid superlattices is studied. The waves income from a fluid or a solid and the transmission medium is a fluid. It is found that the transmission spectrum is closely related to the free-stress surface vibrational modes in the superlattice slab. The correspondence of the transmission peaks with these modes is better when the impedance contrast between the external medium and the superlattice is high. Results for normal and oblique waves in a superlattice of Al/W embedded in Hg, water, or air are presented. As the impedance of the external medium changes it is found new minigaps for the wave transmission inside the allowed frequency region for bulk propagation. Additional structure to the transmission spectrum is produced by localized excitations. Transmission with frequencies in the gap is assisted by the occurrence of surface waves and enhanced transmission is observed inside the frequency bands due to the presence of surface resonances. Theoretical results for a Pt/W superlattice are presented. B.M-M. is grateful to CONACyT Mexico, for a doctorate scholarship. [Work supported by CONACyT Mexico, Grant No. 35541-E.]

11:30

3aPAb8. Directions with negative Poisson's ratio in crystals. Svetlana Tokmakova (N. N. Andreev Acoust. Inst., Shvernic Str. 4, Moscow 117036, Russia)

Materials with negative Poisson's ratio are peculiar in that they expand laterally when stretched. The Poisson's ratio is the ratio of lateral strain in the direction *n* to the longitudinal strain during the stretching of a cylindrical rod along rod axis m. In the paper the stereographic projections of Poisson's ratio for a set of acoustic crystals, metals, alloys, binary compounds, and for the elements of periodic system were computed. From these stereographic projections the cuts with extreme values of Poisson's ratio were determined. Elastic modules of crystals from the manual [Landolt-Bornstein, Numerical Data and Functional Relationships in Science and Technology, New Series, Group III, Vol.18, Berlin (1984)] were applied in calculations. Our evaluations have shown that for most fcc crystals the Poisson's ratio receives negative values for directions of a stretch vector m located inside a cone with an axis [110], relevant to the Euler angles (45,0,90). The parameters of cone were spotted for a set of crystals. For beta cristobalite SiO_2 a decrease of rod cross-section under the uniaxial compression was revealed for the directions of vector m oblique to [111]. Elastic modules of SiO_2 have been taken from Kimizuka *et al.*, Phys. Rev. Lett. **84**, 5548–5551 (2000)]. Some possible applications of materials with a negative Poisson's ratio basing on their unusual acoustic properties are discussed. For the elements of periodic system the results of calculations will also be submitted on the Web-site: www.akin.ru as the color table.

Session 3aSC

Speech Communication: Development, Disorders, Etc. (Poster Session)

Fredericka Bell-Berti, Chair
Department of Speech Communication Sciences & Theatre, St. John's University, 8000 Utopia Parkway, Jamaica, New York 11439

Contributed Papers

All posters will be on display from 8:30 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 10:15 a.m. and contributors of even-numbered papers will be at their posters from 10:15 a.m. to 12:00 noon.

3aSC1. Syntactic location facilitates word segmentation by infants. Elizabeth K. Johnson, Amanda Seidl, and Peter W. Jusczyk (Psych. and Cognit. Sci. Depts., Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218)

English learners begin to segment words and parse speech into prosodic units between 6 and 12 months. Recent experimental evidence suggests that segments are more fully realized at the edges of major syntactic constituents. For example, the [s] in ''bass'' will completely assimilate to [S] in the absence of a major syntactic boundary between the words ''bass'' and ''shoppers.'' In this study, we examine whether the location of words in a phrase (and hence the phonological strength of a word's initial segment) affects infants' segmentation. Eleven-month-olds were familiarized with stories containing the target words ''acorn'' and ''shoppers,'' using the head-turn preference procedure. For half of the infants, ''shoppers'' occurred after a major syntactic boundary, whereas ''acorn'' did not. For the other half, ''acorn'' occurred after a strong syntactic boundary, but ''shoppers'' did not. Infants in the first condition listened significantly longer to isolated versions of ''shoppers'' than to ''acorn.'' In contrast, infants in the second condition listened equally long to both familiar words. These findings suggest that words located at major syntactic boundaries are easier to segment. Our results also provide further evidence that infants' segmentation of vowel-initial words is delayed compared to their segmentation of consonant-initial words.

3aSC2. Infants use of visual information in speech segmentation. George J. Hollich, Peter W. Jusczyk (Johns Hopkins Univ., Baltimore, MD), and Rochelle S. Newman (Univ. of Maryland)

What are infants abilities to segment the running stream of speech in a noisy environment? Infants often find themselves in situations far louder and more complex than the acoustic isolation chambers and minimalist rooms of traditional infant testing. The current series of studies used the headturn preference procedure (with video familiarization) to examine 7.5-month-old infants abilities to use visual/auditory correlations to reliably segment a given speech stream in the face of a distracting voice. Results indicated that infants gained a 10-dB advantage when the visual display matched the acoustic passage. That is, when two blended voices were of equal loudness, infants could use visual correspondences to reliably recognize words presented in the matching video. In contrast, in cases where the target voice was 5 dB softer than the distracting voice or a static picture of the face was presented, infants did not show an ability to segment the speech stream.

3aSC3. Infant-directed vowels are easier to learn for a computer model. Bart de Boer and Patricia K. Kuhl (Ctr. for Mind, Brain, and Learning, Univ. of Washington, Box 357988, Seattle, WA 98195)

When addressing infants, caretakers use a special register, called ''motherese'' or ''parentese.'' Kuhl *et al.* (Science, 1997) demonstrated that the formant patterns of the centers of vowels in this register are farther apart than those in adult-directed speech, and postulated that this helps children learn the categories of speech in their language. In this study, this hypothesis was tested with a computer model. Three words were used: sock, shoe, and sheep. Recordings of these words were available from ten mothers in both infant-directed and adult-directed registers. The first two formants were automatically extracted from the voiced parts of each word. Automatic formant extraction often results in formant patterns with errors, but no corrections were made. After formant extraction, their distributions were compared and a mixture of three Gaussians was fitted to the data points. The results demonstrated that infant-directed tokens not only used a larger part of the acoustic space, but also resulted in a better distribution of the centers of the Gaussians, indicating better patterning of the input. Better patterning of the input makes it easier to learn the categories of speech sounds. [Research supported by the NIH and the Talaris Research Institute.]

3aSC4. Analysis on infant speech with longitudinal recordings. Shigeaki Amano, Tadahisa Kondo, and Sachiyo Kajikawa (NTT Commun. Sci. Labs., NTT Corp., 2-4 Hikari-dai, Seika-cho, Souraku-gun, Kyoto 6190237, Japan)

The utterances of five infants with their parents was recorded every month from their birth until five years old to investigate spoken language development from the viewpoint of acoustic characteristics. Recording time was at least one hour per month. An infant speech database is now being developed from the recordings. Each entry of the database contains a speech file of an utterance and its transcription with some tags such as speaker, utterance category, and clarity. Using a beta version of the database, preliminary analyses were conducted on utterance duration, speaking rate, interutterance interval, and utterance overlap. Some tendencies were observed. For example, utterance duration became longer for infants as a function of infant age, but it was almost constant for parents. On the other hand, the speaking rate was almost constant (about 4 mora/s) for infants but became faster (from 6 to 8 mora/s) for parents as a function of infant age. An interutterance interval of ''Parent to Infant'' correlated with that of ''Infant to Parent,'' even before infants spoke a word. Parent utterance often overlapped infant utterance only after infants spoke two-word sentences. The implication of these tendencies for spoken language development is discussed.

3aSC5. Age-related changes in the acoustic characteristics of voiceless English fricatives. Robert Allen Fox, Shawn Nissen, Julie McGory, and Kimberly Rosenbauer (Speech Percept. and Acoust. Lab., Dept. of Speech and Hearing Sci., The Ohio State Univ., Columbus, OH 43210-1002, fox.2@osu.edu)

This study examines the acoustic characteristics of voiceless fricatives in American English ([f, θ, s, ʃ]) in children and adults (ranging in age from 3 to 70 years). Twenty speakers from seven different age groups produced four different examples of each fricative in word-initial position (in 20 different monosyllabic words varying in terms of medial vowel). Each speaker also produced these fricatives intervocalically (in the context [ɑ Cɑ]). These data are drawn, in part, from a large-scale acoustic database developed in collaboration with ATR. All tokens were recorded using high-quality microphones and direct analog-to-digital conversion to disk at a sampling rate of 22.05 kHz. Acoustic measurements include fricative duration, vowel duration, spectral moments, spectral peak locations, absolute and relative noise amplitude, spectral tilt, and $F2$ onset frequency. Of primary interest is to determine whether these four fricatives can be distinguished using the same subset of acoustic properties as reported in Jongman and colleagues [Jongman *et al.*, J. Acoust. Soc. Am. **108**, 1252–1263] across variations in the age and gender of the speaker. [Work supported, in part, by an INRS Award from NIH and research funding from ATR, Fox, P.I.]

3aSC6. The relation between first graders' reading level and vowel-production variability and presentation format. Kandice Baker (Dept. of Speech, Commun. Sci., and Theatre, St. John's Univ., Jamaica, NY 11439), Anne Fowler (Haskins Labs., New Haven, CT), and Fredericka Bell-Berti (Dept. of Speech, Commun. Sci., and Theatre, St. John's Univ., Jamaica, NY 11439)

The purpose of this research is to determine if children with reading difficulties produce vowels with greater variability than children with normal reading ability. This study examines variability in vowels produced by first grade students as a function of whether they were reading words presented individually in random or blocked format or were naming pictures representing words within a story context. The vowels chosen for this study are /ɪ/, ɛ, and æ, occurring in real monosyllabic words. The specific focus of this research is variability of temporal and spectral aspects of vowel production: vowel duration and $F1$ and $F2$ frequency at vowel onset and at 50 ms intervals through the vowel. The results will be compared with spelling ability (a good measure of reading ability for first graders), to determine if good and poor readers differ in vowel production variability. The results will also be examined to determine if mode of presentation (picture or read word) influences production variability. [Work supported by the U.S. Department of Education, McNair Scholars Program.]

3aSC7. The relation between reading level and vowel production variability in real and nonsense words. Kimberly Lydtin (Dept. of Speech, Commun. Sci., & Theatre, St. John's Univ., Jamaica, NY 11439), Anne Fowler (Haskins Labs., New Haven, CT), and Fredericka Bell-Berti (St. John's Univ., Jamaica, NY 11439)

Past research has indicated possible links between phonological difficulty, speech production variation, and reading problems. This study examines the relations between spelling ability (a measure of reading ability for first graders), on the one hand, and the variability of vowel production, on the other, of poor and good readers in the first grade. The focus of this study was to examine the variability of vowel formant frequencies in real and nonsense words produced in random order by poor and good readers. The vowels chosen for this study are /ɪ/, /ɛ/, and /æ/, occurring in real words (/bVt/) and nonsense (/bVp/) monosyllabic words. Vowel duration and $F1$ and $F2$ frequencies, at vowel onset and at 50-ms intervals through the vowel, will be compared for the poor and good readers (spellers), to determine if the word-status of read items affects production variability. [Work supported by U.S. Dept. of Education, McNair Scholars Program.]

3aSC8. Relationships between acoustic and perceptual measures of stress in nonwords produced by children with developmental apraxia of speech and children with phonological disorders. Benjamin Munson, Elissa Bjorum, and Jennifer Windsor (Dept. of Commun. Disord., Univ. of Minnesota, 115 Shevlin Hall, 164 Pillsbury Dr., S.E., Minneapolis, MN 55455)

This study investigated relationships between acoustic measures of stress in nonwords produced by children with suspected developmental apraxia of speech (sDAS) and children with phonological disorders (PD), and perceptual judgments made by trained listeners. Five children with sDAS and five children with PD produced multiple tokens of eight disyllabic nonwords with iambic and trochaic stress patterns. Measures of vowel duration, $f0$, and intensity were made. Children from both groups produced acoustic differences between stressed and stressless syllables. In a perception experiment, 10 listeners judged whether tokens were produced with stress on the initial or final syllable. Listeners judged that children with sDAS accurately matched target stress contours less often than children with PD. The proportion of listeners judging that a word was produced with final syllable stress was used as the dependant variable in a regression, in which the ratio of vowel duration, $f0$, and intensity between the two syllables served as the independent variables. All three variables predicted a significant proportion of variance in listener judgments, although the overall proportion of variance accounted for was low ($R^2 = 0.212$). Thus, children with sDAS are able to produce contrasts between stressed and stressless syllables, but these are often imperceptible to listeners.

3aSC9. Speech prosody in cerebellar ataxia. Maureen A. Casper (Dept. of Speech and Lang. Pathol., The College of New Rochelle, 29 Castle Pl., New Rochelle, NY 10805), Lawrence Raphael (Adelphi Univ., Long Island, NY), and Katherine S. Harris (Haskins Labs., New Haven, CT 06510)

Ataxic dysarthria has been investigated both acoustically and perceptually without correlation to neurological imaging. The present study sought an acoustic signature with correlation to neurological degeneration for the speech disturbance recognized in cerebellar degeneration. Acoustic measures and magnetic resonance imaging were used to measure speech disturbance in ataxic speakers. The spatial-temporal properties of syllable production in ataxic and normal speakers were examined in twelve speakers (six ataxics and six normals). The speaking task was designed to elicit six different prosodic conditions and four contrastive prosodic events. The normal speakers showed statistically significant differences for all four prosodic contrasts. The cerebellar speakers showed a different pattern of lengthening and syllable reduction from the normal speakers. These differences were found in both syllable duration and $F1$. An estimate of speech deterioration as determined by individual differences between cerebellar and normal subjects' acoustic values of the four prosodic contrasts was then used in correlation analyses with MRI ratings. Moderate correlations between speech deterioration and cerebellar atrophy were found in the measures of syllable duration and f_0. A strong negative correlation was found for $F1$. The normal model described by the prosodic contrasts provided a sensitive index of cerebellar pathology with quantitative acoustic analyses.

3aSC10. Coarticulation in dysarthria. Kris Tjaden, Alexa Kozak, and Jessica Kleinhaut (Dept. of Commun. Disord. & Sci., The State Univ. of New York at Buffalo, 3435 Main St., Buffalo, NY 14214-3005)

The dysarthrias are a group of speech disorders resulting from impairment to nervous system structures important for the motor execution of speech. Numerous studies have examined how dysarthria impacts articulatory movements and the resulting changes in vocal tract shape. Few studies of dysarthria, however, consider that articulatory events and their acoustic consequences overlap or are coarticulated in connected speech,

although coarticulation has been studied extensively in neurologically normal speakers and figures prominently in normal speech production theories. In so far as coarticulation presumably would have a similar status in a speech production theory of dysarthria, further investigation of coarticulatory patterns in dysarthria appears warranted. The current study will examine intersyllabic, anticipatory vowel-to-vowel coarticulation in normal speakers, and speakers with Multiple Sclerosis, a disease characterized by plaques in the white matter of the CNS. The speech sample consists of 3 repetitions of 16 target words in the phrase ''It's a _____ again.'' Phrases were produced in normal, loud, and slow speaking conditions. The effect of target word vowels on /ə/ in the word ''a'' will be inferred from $F2$ measures. Within conditions, coarticulatory differences will be compared for MS and Control speakers. Coarticulatory differences across speaking conditions also will be examined.

3aSC11. The correlation of vowel space and speech intelligibility for individuals with cerebral palsy. Huei-Mei Liu, Feng-Ming Tsao, and Patricia K. Kuhl (Ctr. for Mind, Brain, and Learning, Univ. of Washington, Seattle, WA 98195)

The relationship between the area of the vowel space ($F1/F2$ space) and speech intelligibility was studied in a group of 20 Mandarin-Chinese-speaking young adults with cerebral palsy. Subjects read aloud 18 bisyllabic words containing the vowels /i/, /a/, and /u/ using their habitual speaking rate. The words were identified by three normal listeners and the percent corrects of vowel identifications were used as speech intelligibility for the individual cerebral-palsied speakers. Results revealed that cerebral-palsied speakers' intelligibility varied greatly across individuals (scores ranged from 36% to 100%). Speakers with cerebral palsy exhibited smaller vowel space areas when compared to 10 age-matched controls. The correlation between vowel space area and speech intelligibility was significant ($r=0.612$, $p<0.005$). Cerebral-palsied speakers with larger vowels spaces were generally more intelligible than speakers with reduced vowel spaces. These results suggest that vowel space area can serve as an important component of overall estimates of speech intelligibility, and will be related to data collected in studies comparing infant-directed (ID) versus adult-directed (AD) speech, which show that speakers use significantly larger vowel areas when engaged in ID speech. [Work funded by NIH, HFSP, and the Talaris Research Institute.]

3aSC12. Binaural dichotic presentation to reduce the effect of increased temporal and spectral masking in sensorineural hearing impairment. Prem C. Pandey, Dakshayani S. Jangamashetti, and Alice N. Cheeran (EE Dept., IIT Bombay, Powai Mumbai 400076, India, pcpandey@ee.iitb.ac.in)

Sensorineural hearing loss is characterized by increased temporal and spectral masking, resulting in degraded speech perception. Earlier investigations have shown that binaural dichotic presentation using comb filters with complementary magnitude responses improved speech perception, particularly the place feature. Further, inter-aural switching with periodic trapezoidal fading functions improved the perception of duration feature. For simultaneously reducing the effect of the two types of masking, a scheme has been devised with a pair of time-varying comb filters, with bands corresponding to auditory critical bands. Thus the spectral components in neighboring critical bands do not mask each other, and the sensory cells on the basilar membrane get a relaxation time due to sweeping of the filter passbands. The scheme was implemented using linear phase 256-coefficient FIR filters. Magnitude responses were cyclically swept with 20-ms period, in steps by selecting the filter coefficients for each step from a pre-calculated set. Listening tests involved closed set identification of 12 vowel–consonant–vowel syllables. The processing scheme resulted in the improvement of recognition scores, response time, and transmission of features, particularly place and duration, indicating reduction in the effect of spectral and temporal masking.

3aSC13. Comb filters for binaural dichotic presentation to improve speech perception by persons with bilateral sensorineural hearing impairment. Alice N. Cheeran, Prem C. Pandey, and Dakshayani S. Jangamashetti (EE Dept., IIT Bombay, Powai Mumbai 400 076, India, pcpandey@ee.iitb.ac.in)

Intelligibility of speech for persons with sensorineural hearing impairment gets reduced due to increased spectral masking. For persons with bilateral loss, improved speech perception has been reported with binaural dichotic presentation using comb filters having complementary magnitude responses, based on auditory critical bands. In this scheme, spectral components in adjacent critical bands do not mask each other as they get presented to different ears. The objective here was to design a set of comb filters with transition crossovers adjusted for a low perceived spectral distortion. For this, a slowly swept tone presented through the two comb filters should not have a change in its perceived loudness at crossover frequencies. Experiments showed that the change in loudness was negligible with crossover response lying between -4 dB and -6 dB with respect to passband response. With this constraint for response at the crossover frequencies, comb filters with responses corresponding to 18 critical bands were designed as 256-coefficient linear phase FIR filters, using frequency sampling technique. The filters have pass band ripple of 1 dB and stop band attenuation of 30 dB. These filters have resulted in improvement in speech quality and speech perception.

3aSC14. Training a vowel contrast using visual articulatory feedback. Richard S. McGowan (CReSS LLC, 1 Seaborn Pl., Lexington, MA 02420), Linda Ferrier (Northeastern Univ., Boston, MA), Karen Chenausky, and Joel MacAuslan (S-T-A-R Corp., 4 Militia Dr., Lexington, MA 02421)

The /i/–/I/ distinction can be difficult to make in speech for deaf, dysarthric, and many non-native English speakers. A computer program was designed that prompted speakers, with orthograpraphy, to produce a simple single syllable word containing either vowel and recorded the speakers response. The program provided feedback in one of two forms: either auditory (AF) or auditory and visual articulatory feedback (VAF). The auditory feedback was the actual and correct pronunciation of the word. VAF provided schematic representations of where the tongue was according the speaker's actual production of the vowel in relation to where it would be placed for the best production of the target vowel. Articulatory representations were inferred automatically based on the measured formant frequencies of the speaker's production. Four subjects who were particularly poor at making the vowel distinction were recruited to test VAF in a single subject case study design. After baseline scores for the vowel distinction were established in four sessions, each subject was given training using each of AF and VAF, with no words the same for each of the types of training. There was more improvement in the vowel distinction for the auditory and visual feedback compared to the auditory feedback alone.

3aSC15. Gesture-speech interaction in the SmartKom project. Antje Schweitzer (Inst. for Natural Lang. Processing, Univ. of Stuttgart, Germany, antje.schweitzer@ims.uni-stuttgart.de)

The goal of the SmartKom project is to create an intuitive multi-modal dialog system. Information is presented on the display by an artificial lifelike character, combining the output modalities speech, mimics, and gesture. Multi-modality poses new requirements for speech synthesis.

Speech may be accompanied by gestures, adding three new aspects: First, synchronization of lip movements; second, temporal alignment of speech and gesture; and third, effects of gestures on prosody. This paper focuses on the last two points. Building on de Ruiter's (2000) Sketch Model, speech is synthesized independently from the temporal structure of the accompanying gesture. Preparation phase and retraction phase of gestures can be adjusted to align the relevant speech material with the stroke phase. Concerning prosody, results of a pilot study on intonation of deictic elements show that speech material accompanied by gestures is more likely to be prosodically prominent. Gestures in the SmartKom system are always triggered by linguistic properties of the accompanying speech. Most of the gestures are pointing gestures, which are triggered by deictic elements. Such linguistic information is integrated into the concept input to the synthesis module and is taken into account during prosody generation.

3aSC16. Auditory-visual speech perception examined by functional MRI and reaction time. Kaoru Sekiyama (Future Univ. Hakodate, Kameda Nakano, Hakodate 041-8655, Japan, sekiyama@fun.ac.jp) and Yoichi Sugita (Adv. Inst. of Sci. and Technol., Tsukuba, Japan)

In face-to-face communication, auditory and visual speech cues are easily integrated as demonstrated by the McGurk effect. The integration processes are investigated by measuring brain activity (*f*MRI) and reaction time. The subjects were 10 native speakers of Japanese. The stimuli were /ba/, /da/, and /ga/ uttered by three female talkers. The audiovisual stimuli were the McGurk type stimuli consisting of discrepant auditory and visual syllables (e.g., audio /ba/ was combined with video /da/ or /ga/). We compared brain activity during audiovisual speech perception for two sets of conditions differing in the intelligibility of auditory component of speech. In each condition, the subjects were asked to identify spoken syllables. When the auditory speech was intelligible, a brain area for visual motion processing was quiet, whereas the same visual area was active when speech was harder to hear. Thus visual information of the mouth movements was processed more intensively when speech was harder to hear. Reaction time was faster for the low-intelligibility condition than for the high-intelligibility condition, suggesting a top-down suppression of visual processing when auditory speech was intelligible. The integration processes seem to involve a process to actively find the optimal weight of different modalities under given circumstances.

3aSC17. Objective evaluation of nasality using laser Doppler vibrometer. Masanobu Kumada (Natl. Rehabilitation Ctr. for the Disabled, 4-1 Namiki, Tokorozawa 359-8555, Japan, kumada@rehab.go.jp), Toshiaki Kaneko (Hitachi Ltd., Tokyo, Japan), Takahiko Ono, Masanao Ohashi, Koji Ishida (Onno Sokki, Yokohama, Japan), Kazuo Nakayama (Univ. of Yamagata, Yamagata, Japan), Noriko Kobayashi, Hajime Hirose (Kitazato Univ., Sagamihara, Japan), Niro Tayama (Univ. of Tokyo, Tokyo, Japan), and Seiji Niimi (Intl. Univ. of Health and Welfare, Otawara, Japan)

Objective measurement of nasality is very important in terms of clinical evaluation. But there is no definite methodology of objective evaluation of nasality. We already reported that new methodologies of the Laser Doppler Vibrometer (LDV) and Sub-miniature Electret Microphone (SEM), both of which could detect the nasal/non-nasal difference of sustained phonemes, seemed valid to evaluate nasality [Kumada *et al.*, 140th ASA Meeting (2000); Kumada *et al.*, 141st ASA Meeting (2001)]. LDV, used recently in an industrial field as well as medical/biological field, is the generator and the detector of the laser to detect the velocity of the object's vibration by Doppler's effect. LDV seemed a powerful tool for speech physiology by which vibration of a body wall can be measured during speech and singing, without any contact receiver. The objectives were to evaluate the degree of nasality of phonemes in words by LDV and SEM, and to discuss the possibility of the clinical use of these methodologies. The subjects were 5 healthy Japanese males (Tokyo dialect speakers). The materials were Japanese words /shinkansen/ and /shimbunshi/. Both LDV and SEM systems could detect a nasal/non-nasal difference in spoken words, and seemed valid to evaluate nasality in speech. Further study is planned using patients (e.g., cleft-palate) as well as healthy subjects.

3aSC18. Analysis of yes–no and confidence-rating word recognition response formats. Edward L. Goshorn and Jennifer D. Goshorn (Speech Dept., Louisiana Tech. Univ., 3165 Tech Station, Ruston, LA 71272)

The effects of two response formats (yes–no and confidence-rating) on percent correct, d', and response criterion were examined. Diagnostic Rhyme Test words were presented in white noise (3 and 6 dB signal to noise ratios) to eight normal hearing subjects at 70 and 90 dB SPL; three replicates were obtained. For the yes–no format, subjects checked ''yes'' or ''no'' that a single response option matched the stimulus. For the confidence-rating format, subjects checked 1 of 2 response options and indicated confidence, or lack thereof, that the response matched the stimulus by marking ''yes'' or ''no.'' Both formats yield signal detection theory (SDT) estimates of d' and response criterion, but differ in categorizing responses as hits, misses, false alarms, and correct rejections. A significant difference was found between response formats for percent correct and response criterion but not for d'. The effects of signal to noise ratio were significant but parallel across response format for percent correct, d', and response criterion. Results indicate that response format may significantly affect conventional measures such as percent correct but not SDT measures that separate sensitivity and response criterion. The use of SDT measures in word recognition procedures is supported.

3aSC19. The relationship between auditory temporal resolution and auditory object formation. Camille C. Dunn and Thomas D. Carrell (Commun. Disord., Univ. of Nebraska, Lincoln, NE 68583)

The goal of the present experiment was to determine whether a relationship exists between auditory temporal resolution and the ability to use amplitude modulation for auditory object formation in sentence perception. A standard gap detection task was used to evaluate temporal resolution in listeners [D. P. Phillips, T. L. Taylor, S. E. Hall, M. M. Carr, and J. E. Mossop, J. Acoust. Soc. Am. **101**, 3694–3705 (1997)]. Each stimulus consisted of a leading element, a silent duration, and a trailing element. Both within- and between-channel detection was measured. An adaptive tracking procedure (PEST) was used to determine the 70%-correct gap durations for each subject. Next, a sentence identification task was used to evaluate listeners' abilities to use amplitude modulation to extract speech from noise. Time Varying Sinusoidal (TVS) sentences were created for this purpose [T. D. Carrell and J. M. Opie, Percept. Psychophys. **52**, 437–445 (1992)]. Half of the TVS sentences were amplitude modulated at 100 Hz and half remained unmodulated. The difference in intelligibility between the modulated and unmodulated sentences was used as a measure of auditory object formation. No correlation was found between gap discriminability and auditory object improvement among the subjects we tested. A wider range of subjects is currently being investigated.

Session 3aSP

Signal Processing in Acoustics, Biomedical Ultrasound/Bioresponse to Vibration, Acoustical Oceanography and Underwater Acoustics: Theory and Applications of Acoustic Time Reversal

David H. Chambers, Chair
Lawrence Livermore National Laboratories, P.O. Box 808, L-154, Livermore, California 94551

Chair's Introduction—8:25

Contributed Papers

8:30

3aSP1. Time reversal mirror in a reverberant structural network. Joe Dickey (Ctr. for Nondestruct. Eval., Johns Hopkins Univ., 3400 N. Charles St., MD 21218, dickey@jhu.edu) and John Ertel (U.S. Naval Acad., Annapolis, MD 21402-5026)

The time domain response of a reverberant structural network of connected 1-D wave bearing systems excited by a spatial/temporal impulse is calculated and displayed in a spatial/temporal format. The response of a sensor at an arbitrary point in the network will be a time series of events. A segment of this time series is recorded, reversed in time, and used as an input to the Green function model of the network at the point where it was recorded. The model then shows the constructive interference at the original source location; however, it is not possible to ''run the movie backward'' in any strict sense. Constructive interference occurs at other alias locations, and the ability to locate the source degrades with increased damping and complexity. The advantage of using structural intensity and nonphysical modeling ''tricks'' such as negative damping (do not try this experimentally) and altered reflection coefficients after the injection of the time-reversed pulse are demonstrated. This technique can be useful in locating impact points in structural networks.

8:45

3aSP2. Experimental confirmation of time reverse operator eigenvalues on target scattering moments. Alan W. Meyer and D. H. Chambers (Lawrence Livermore Natl. Lab., P.O. Box 808, L-154, Livermore, CA 94551)

Laboratory experiments confirm multiple distinct, scatterer dependent, eigenvalues of the time reverse operator (TRO) for a linear time reverse mirror (TRM). Prada and Fink [Wave Motion **20**, 151–163 (1994)] discuss a one-to-one correspondence between the eigenvalues of the TRO and each scatterer. Chambers and Gautesen [J. Acoust. Soc. Am. **109**, 2616–2624 (2001)] extend this theory to equate the TRO eigenstates to functions of the induced scattering moments of the target object. Each scatterer has a monopole moment associated with the compressibility contrast between the object and the surrounding medium. Likewise, a density contrast will induce multiple dipole moments. Each of these moments results in an eigenstate of the TRO. Experimentally the TRO was determined for two cylindrical scatters in two different media: glass and aluminum. Procedurally, in each case the eigenstates for the TRO were first determined. The DORT method was then applied and the focal point determined. An association between the scattering moment and an eigenstate is confirmed when more than one eigenstate focused the TRM to the same scatter. [This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-ENG_48.]

9:00

3aSP3. Spatial and frequency selectivity of a matched filter array in a conference room. Deborah M. Grove and Joseph C. French (Ctr. for Adv. Information Processing, Rutgers Univ., 96 Frelinghuysen Rd., Piscataway, NJ 08854, dgrove@caip.rutgers.edu)

Matched filter array (MFA) processing uses impulse responses (IRs) in its algorithm to provide high signal-to-noise and 3-D spatial selectivity. The spatial and frequency selectivity of the MFA is shown for an array of eight randomly placed microphones in a 0.7-s reverberant conference room. The IRs were measured using nine averages of a 16 383 sample maximum length sequence at a 16-kHz sample rate from a loudspeaker representing on- and off-focus conditions. The MFA response was formed offline by convolution of the time-reversed IR with an IR representing the incoming source position to each microphone, with a final summation over the array. Truncation of the impulse responses is applied to remove anticausal echo. The spatial selectivity of this array is on the order of 25 cm from the focus. The influence of the room transfer function on the frequency response of the MFA is shown. It is seen that the frequency focal volume immediately degrades off-focus for the high frequencies first and that the off-focus condition causes a time-delay in the array response. The influence of individual matched filters versus the array response on the spatial selectivity is explained. [Work supported by Intel and Texas Instruments.]

9:15

3aSP4. Transmission mode acoustic time reversal imaging with application to geophysical vertical seismic profiling. Sean K. Lehman (Lawrence Livermore Natl. Lab., L-154, 7000 East Ave., Livermore, CA 94550, lehman2@llnl.gov) and Anthony J. Devaney (Northeastern Univ., ECE, 360 Huntington Ave., Boston, MA 02115)

We develop the theory of transmission mode time reversal imaging for arbitrary transducer geometry. Although it is not possible to implement such a system in practice, we prove that transmission mode time reversal imaging is achievable *a posteriori* from the scattered field time series as measured at the receivers. The theory is developed for any arbitrary geometry in which the transmitters and receivers are not colocated. We provide simulated proof-of-principle and finite-difference time-domain (FDTD) examples designed to model the case of geophysical vertical seismic profiling in which a series of sources are excited on the surface and the scattered fields are measured on a chain of receivers in a bore hole.

9:30

3aSP5. Application of time reverse acoustics to acoustic measurements in a noisy environment. Sewon Park and Gerald C. Lauchle (Grad. Prog. in Acoust., Penn State Univ., State College, PA 16803)

For acoustic measurements made within noisy environments, time reversed acoustic methods can be used to improve the measurement. After reversing the array received time signal, rebroadcast of the reversed sig-

nals recreates the original source wave field at its original source location. In some measurement cases a background noise is created in the process of generating the source signal of interest; for example, wind and water tunnel acoustic experiments. By using time reverse acoustics, the signal-to-noise ratio can be improved. First, the source signal together with the background noise is recorded using an array of many reciprocal transducers. After the tunnel flow is stopped, the recorded signals are played in a time-reversed fashion through the reciprocal transducer elements of the array. If a receiver is located at the correct source location, only the source part of the rebroadcast signal is superimposed positively, with negligible influence from the background sources that originate elsewhere. Here, we describe noisy environment, free-field, time reversed acoustic simulations. The computer simulations show an improved signal-to-noise ratio. The advantages and disadvantages of using continuous or pulsed recorded signal are discussed. [Work supported by the Applied Research Laboratory E/F Program, and ONR, Code 333.]

9:45

3aSP6. Laboratory experiments of time reversed acoustics applied to sonar in shallow water waveguides. Mitchell N. Shipley (Appl. Res. Lab., Penn State Univ., State College, PA 16804), John Stokely, Kevin B. Smith, and Andres Larraza (Naval Postgrad. School, Monterey, CA 93943)

Active sonar research at the Naval Postgraduate School Advanced Acoustic Research Laboratory (AARL) has focused on real time adaptation of the active sonar transmit signal. Recent work with Time Reversal Acoustics (TRA) has demonstrated the ability of this approach to adapt in real time to complex environments with little knowledge of the environmental parameters. The present work applies TRA techniques to shallow water active sonar to improve target echo SNR. The approach is to conduct scaled laboratory tank experiments of the TRA sonar and compare results to standard matched filter active sonar processing techniques. Results of laboratory TRA sonar experiments will be presented, to include the high SNR case, the effects of increasing noise levels, and the TRA sonar approach as applied to an active barrier concept.

10:00–10:15 Break

10:15

3aSP7. Preliminary results of water tunnel acoustic measurements using a time reversal mirror. Christopher Barber (Naval Surface Warfare Ctr., Carderock Div., 9500 MacArthur Blvd., West Bethesda, MD 20817-5700), Gerald C. Lauchle (Grad. Prog. in Acoust. and Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804), and Nelson J. Kottke (Grad. Prog. in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804)

A concept demonstration test was recently conducted at the Garfield Thomas Water Tunnel at the Penn State University Applied Research Laboratory using a time reversal mirror to acquire acoustic measurements of a cavitation noise source. The objective of the test was to demonstrate improved acoustic measurement capabilities compared to conventional underwater acoustic measurement techniques. A preliminary analysis of the data indicates an increased signal-to-noise ratio resulting from the time-reverse playback of recorded data compared to real-time measurements made using either omnidirectional or sum-beam sensors. Experimental results are presented, including a novel method for source localization based on the difference between the time-forward and time-reverse playback of recorded data. [Work supported by ONR Code 334, Naval Surface Warfare Center Carderock Division, and the Applied Research Laboratory.]

10:30

3aSP8. Robust time-reversal focusing in a random ocean channel. Seongil Kim, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, G. F. Edelmann (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238), Tuncay Akal (SACLANT Undersea Res. Ctr., 19138 La Spezia, Italy), Rick P. Millane (Purdue Univ., West Lafayette, IN 47907-1160), and Daniela DiIorio (Univ. of Georgia, Athens, GA 30602)

Since time reversal focusing utilizes the reciprocity of a static medium, its performance can be degraded in a nonstatic medium, where the propagation conditions change during the time of forward and backward propagations. In this study, a method for robust time-reversal focusing is investigated based on a method developed for matched-field processing [Krolik, J. Acoust. Soc. Am. **92**, 1408–1419 (1992)]. Instead of using a single probe source pulse, the robust method employs multiple probe source signals obtained over a period of time where each ping represents a different propagation condition. A singular value decomposition of the signal matrix of the received probe source pulses shows that over 90% of the acoustic energy is represented by the first few singular values, suggesting that the effective dimension of the pressure field resulting from the sound speed fluctuations is small. The back-propagated field weighted by the linear combination of the dominant singular vectors produces stable focusing for longer time than that with a single probe pulse. The method is useful in nonstatic propagation conditions and in a situation where frequent probe signals are not available. [Work supported by ONR.]

10:45

3aSP9. Shallow water performance of a moving time-reversing array. Karim G. Sabra and David R. Dowling (Dept. of Mech. Eng., Univ. of Michigan, Ann Arbor, MI 48109-2121)

Future active sonar and underwater communication systems for use in unknown shallow ocean waters may be developed from the automatic spatial and temporal focusing properties of time-reversing arrays (TRAs). Previous simulation results suggest that stationary TRAs will work well with moving sources when the source Mach number is low. This presentation describes simulation results for the case when the TRA is moving and the source is stationary. When the array is moving, the forward and backward propagation necessary for TRA operations is influenced by the Doppler effect in an asymmetric manner. Here, numerical results for the characteristics of the retrofocused field will be presented for a range-independent sound channel based on wide-angle parabolic-equation (RAM by Dr. M. Collins of NRL) or wave number integration (OASES by Professor H. Schmidt of MIT) calculations. Different combinations of array motion and geometry for acoustic center frequencies of several hundred Hertz and nominal ranges up to 10 km will be included, as well as the influence of random bottom roughness. Issues concerning the implementation of a moving TRA in a realistic environment will also be addressed. [Work sponsored by ONR.]

11:00

3aSP10. A model-based acoustic time-reversal mirror for robust variable focusing. Jean-Pierre Hermand (Dept. of Optics and Acoust., Université Libre de Bruxelles, av. F.-D. Roosevelt 50, CP 194/5, B-1050 Brussels, Belgium)

A natural extension of the acoustic time-reversal mirror (TRM) is investigated where *predicted* receive array signals are backpropagated to an arbitrary *virtual* source location. First, large time-bandwidth product (TB), broadband signals propagating from a *real* probe source to the array are processed iteratively by a model-based matched filter (MBMF) receiver in searching for the environmental parameters that best correspond to the measured acoustic-impulse responses. The process ends when most of the time-spread energy across the array is recombined *coherently* (processing gain). Second, the resulting acoustic-channel model is used to synthesize the transmit signals that will focus their energy at the desired range and depth (the virtual source). The procedure overcomes the noise and reverberation limitations inherent to the TRM operation. The at-sea, focusing performances of a model-based TRM were predicted from related MBMF experiments: range-depth localization of a distant towed source in a duct

(WEST SARDINIA 89–90) and bottom geoacoustic characterization in shallow water (YELLOW SHARK 94–95) [Hermand *et al.*, IEEE J. Oceanic Eng. **18**, 447–465 (1993); **24**, 41–66 (1999)]. Simulation results using YELLOW SHARK 94, large TB, broadband vertical array data will be compared with TRM 96–97 experimental results [Hodgkiss *et al.*, J. Acoust. Soc. Am. **105**, 1597–1604 (1999)] obtained along the same transect.

11:15

3aSP11. Decision-directed passive phase conjugation: A robust and simple architecture for array receivers in the underwater communications channel. John A. Flynn, James A. Ritcey (Dept. of Elec. Eng., Univ. of Washington, Box 352500, Seattle, WA 98105), and Warren L. J. Fox (Univ. of Washington, Seattle, WA 98105)

A decision-directed extension to the passive phase conjugation (PPC) technique of array demodulation for underwater communications channels is presented. The PPC method utilizes a probe signal from the transmitter to exploit the time-reversal concept in a one-way fashion at the receiver on subsequent unknown data signals [D. R. Jackson *et al.*, Conf. Rec. 34th Asilomar Conf. on Signals, Syst., and Comp., Vol. 1, 680–683 (2000)]. PPC demonstrated excellent equalization of communication bursts in shallow water with sparse array reception. The decision-directed extension to PPC presented here virtually eliminates the need for repeated channel probing, while improving channel estimation in rapidly varying, noisy environments. The technique utilizes decision-directed channel modeling in a fixed-window linear statistical framework, a deterministic least squares optimization criteria, and robust conjugate-gradient numerical methods driving the PPC demodulator. The resulting architecture scales linearly with the number of sensors, has a per-sensor complexity competing with fast-RLS algorithms, and avoids performance limitations of both LMS and E-RLS. Three difficult shallow water demonstrations at 1–5 km show robust channel tracking with good BER and excellent equalization performance. [Work supported by ONR.]

11:30

3aSP12. Detection based on multiple time-reversal guide-sources. Charles F. Gaumond (Acoust. Div., Naval Res. Lab., Washington, DC 20375-5320)

The problem of inclusion of multiple guide-sources in a time-reversal detection system is discussed. The objective of the detector is the coverage of a given volume over range and depth in a known environment. This detector should rely on as few guide-sources as possible. This detection problem is analyzed in terms of a separable kernel receiver (SKR). The performance of the SKR in terms of the number and placement of guide-sources is shown to depend on the 2-D inner-product of the guide-source covariance with the expected signal covariance. This single number can be used as an overall measure of system quality. Numerical examples are shown for several shallow water environments with various numbers and placements of guide-sources. The use of extended reflectors for use in an active system is also considered. [Work supported by ONR.]

WEDNESDAY MORNING, 5 DECEMBER 2001 ROOM 304, 8:30 TO 11:45 A.M.

Session 3aUW

Underwater Acoustics, Acoustical Oceanography and Animal Bioacoustics: Ocean Research and Marine Mammal Regulatory Issues

David L. Bradley, Chair
Applied Research Laboratory, Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16802

Chair's Introduction—8:30

Invited Papers

8:35

3aUW1. Marine Mammal Protection Act (MMPA), Endangered Species Act (ESA), and regulations for Scientific Research Permits. Gene Nitta and Ann Terbush (Office of Protected Resources, Div. of Marine Mammal Permits, Conservation, and Education, Silver Spring, MD 20910, gene.nitta@noaa.gov)

The MMPA, ESA, and Fur Seal Act, mandate the protection and conservation of marine mammals and other protected species, and prohibit the taking, importation, and export of protected species. Exceptions for scientific research, enhancement, public display, and commercial and educational photography, are allowed under special exception permits or other appropriate authorizations issued under Section 10 of the ESA and Section 104 of the MMPA. Permit issuance must also meet NEPA, ESA Section 7, CITES, and where necessary, CZMA requirements. NOAA Fisheries develops and implements policies and regulations for issuance of permits and authorizations to take marine mammals for the activities noted above, and organizes and executes the national program for tracking captive marine mammals. Marine mammal permits are coordinated with NOAA Regional Offices, the Marine Mammal Commission, the Fish and Wildlife Service, and the Animal and Plant Health Inspection Service. Input on permit applications is received from other Federal, State, and local agencies, the academic community, public display institutions, nongovernmental organizations, and the general public. NOAA fisheries coordinates national policy for minimizing harassment from human recreational activities directed at marine mammals and conducts outreach to the public about the permit program and responsible wildlife viewing practices.

9:00

3aUW2. Marine Mammal Protection Act (MMPA) and Endangered Species Act (ESA) regulations for marine mammal nonfishery interactions. Simona Perry Roberts and Donna S. Wieting (Office of Protected Resources, Marine Mammal Conservation Div., NOAA Fisheries, 1315 East–West Hwy., Silver Spring, MD 20910, simona.roberts@noaa.gov)

The Office of Protected Resources Marine Mammal Conservation Division develops, implements, and administers programs for the protection, conservation, and recovery of marine mammals under the Marine Mammal Protection Act (MMPA) and the Endangered Species Act (ESA). The Division's nonfishery interaction program is tasked with implementation of Sec. 101(a)(5) (A–D) of the MMPA, which provides two types of authorizations for allowing the incidental, but not intentional, taking of small numbers of marine mammals by U.S. citizens engaged in a specified, nonfishery related activity. The first type of authorization, the Letter of Authorization (LOA), requires that regulations be promulgated. These LOA regulations may be effective for a 5 year period, with a new letter, containing specific mitigation and monitoring requirements, issued on an annual basis. The second type of authorization, the Incidental Harassment Authorization (IHA), does not require promulgation of regulations but does require a 30-day public comment period. The IHA is valid for 1 year only. The estimated time for processing LOA and IHA requests ranges from 1 year to 120 days; however, depending on other requirements, this time period may be lengthened. NMFS makes its determination to issue or deny requests based on three criteria: negligible impact, small numbers, and unmitigable adverse impacts.

9:25

3aUW3. Reauthorizing and amending the MMPA and ESA: Issues, status, and prospects in the 107th Congress. Eugene Buck (Resources, Sci. and Industry Div., Congressional Res. Service, 101 Independence Ave., S.E., Washington, DC 20540-7450)

The CRS has provided specialized research and policy analysis to the Congress for more than 80 years. As a legislative branch agency, the CRS works exclusively for Members of Congress, Committees, and Congressional staff on a confidential, nonpartisan basis. The CRS analytical staff are internationally-recognized experts who apply multidisciplinary expertise at all stages of the legislative process to assist Congress in identifying, analyzing, and formulating legislative proposals as well as responding to constituent needs. The 107th Congress is considering whether to reauthorize and amend the Endangered Species Act of 1973 (ESA). Major issues in recent years have focused on whether to incorporate further protection for property owners and reduce regulatory impacts, or whether to strengthen species protections through a variety of procedural changes. The authorization of appropriations for the Marine Mammal Protection Act (MMPA) expired at the end of FY1999. The issue for the 107th Congress is whether, and if so under what conditions, to reauthorize and amend this Act to reflect and address the concerns of various interest groups. For both the ESA and MMPA, issues capturing congressional attention will be identified, the current status of legislation to address these issues will be summarized, and the prospects for enactment will be discussed.

9:50–10:05 Break

10:05

3aUW4. Coastal zone management regulations in California and other coastal states. Mark Delaplaine (California Coastal Commission, 45 Fremont St., Ste. 2000, San Francisco, CA 94114)

The California Coastal Commission is California's designated coastal management agencies for the purpose of administering the federal Coastal Zone Management Act (CZMA) over Pacific Ocean waters offshore of California. Once a state's coastal management program is certified by the federal government, the CZMA gives state coastal management agencies regulatory control (federal consistency review authority) over all federal agency activities and all federally licensed, permitted or funded activities (regardless of the location of the activity), if the activity affects the land or water uses or natural resources of the coastal zone. The regulations and the regulatory process under the CZMA (and under State law for activities in state waters) in California will be discussed, with an emphasis on the use of underwater acoustic devices (including seismic surveys, military sonar, and acoustic thermometry, and other research) and the evolution of policy development over the past 5–6 years. Comparisons and contrasts with procedural and policy positions taken by other states under the CZMA will also be made.

10:30

3aUW5. Environmental compliance in naval research. John Quinn (Office of Naval Res., 800 N. Quincy St., Arlington, VA 22217)

The Office of Naval Research (ONR) promotes, coordinates, and executes the science and technology programs of the Department of the Navy. Acting principally through universities, government laboratories, and nonprofit organizations, the ONR sponsors basic and applied research, and advanced technology development, in ocean acoustics and many other disciplines. In these efforts the ONR and its associates comply with a host of environmental statutes and executive orders, and with Navy policy regarding marine environmental protection. This paper reviews these requirements as they apply to the ONR's ocean acoustic efforts, and the processes through which the ONR carries out its environmental responsibilities.

10:55

3aUW6. Regulating underwater sound in Eastern Canada: An operational perspective from the Department of Fisheries and Oceans. Paul A. Macnab (Oceans and Coastal Management Div., Oceans and Environment Branch, Fisheries and Oceans Canada, Bedford Inst. of Oceanogr., P.O. Box 1006, Dartmouth, NS B2Y 4A2, Canada)

Fisheries and Oceans in Canada are responsible for policies and programs that support economic, ecological and scientific interests in Canada's marine waters. Central to the Department's mandate is the conservation of living marine resources, including mammals. Canadian regulations, permits and approval processes will be discussed with an emphasis on acoustic devices and related noise issues

on the Scotian Shelf. Commercial seismic research on the Shelf has added considerably to the sound dependent marine research typically undertaken by government and naval scientists. This increase in noise has prompted calls for cumulative environmental assessments that address additive and synergistic effects. Existing mitigation measures include sensitive area avoidance and coordination to avoid spatial and temporal overlap. Operations are restricted near the Sable Gully, a large canyon recognized as the most significant cetacean habitat on the Scotian Shelf. Evolving nonregulatory approaches include codes of conduct, precautionary buffer zones and voluntary compliance with operational guidelines. Work conducted by Defence Research Establishment Atlantic on historical levels of ambient noise has been especially instructive in these matters. Additional measurement and modeling expertise is needed to help establish safe operating distances and environmental quality standards that could be applied to all anthropogenic sound sources near the Sable Gully.

11:20–11:45
Panel Discussion

WEDNESDAY MORNING, 5 DECEMBER 2001 ROOM 209, 8:30 TO 10:30 A.M.

Meeting of Accredited Standards Committee (ASC) S2 Mechanical Vibration and Shock

to be held jointly with the

U.S. Technical Advisory Group (TAG) Meeting for: ISO/TC 108 Mechanical Vibration and Shock ISO/TC 108/SC1 Balancing, including balancing machines ISO/TC 108/SC2 Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures ISO/TC 108/SC3 Use and calibration of vibration and shock measuring instruments ISO/TC 108/SC5 Condition monitoring and diagnostics of machines and ISO/TC 108/SC6 Vibration and shock generating systems

R. J. Peppin, Chair S2
5012 Macon Road, Rockville, Maryland 20852

D. J. Evans, Vice Chair S2 and Chair of the U.S. Technical Advisory Group (TAG) for ISO/TC 108, Mechnical Vibration and Shock and Chair of the U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 3 Use and calibration of vibration and shock measuring devices
National Institute of Standards and Technology (NIST), 100 Bureau Drive, Stop 8221, Gaithersburg, Maryland 20899-8221

M. Schonfeld, Chair of the U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 1 Balancing, including balancing machines
Schenck Trebel, 535 Acorn Street, Deer Park, New York 11729

A. F. Kilcullen, Chair of the U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 2 Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures
5012 Woods Road, Hedgesville, West Virginia 25427

R. Eshleman, Vice Chair of the U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 2 and Chair of the U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 5 Condition monitoring and diagnostics of machines
Vibration Institute, 6262 Kingery Highway, Ste. 212, Willowbrook, Illinois 60514

Accredited Standards Committee S2 on Mechanical Vibration and Shock. Working group chairs will report on the status of various shock and vibration standards currently under development. Consideration will be given to new standards that might be needed over the next few years. There will be a report on the interface of S2 activities with those of ISO/TC 108 and its subcommittees including plans for future meetings of ISO/TC and/or its Subcommittees. The Technical Advisory Group for ISO/TC 108 and the subcommittees listed above consists of members of S2 and other persons not necessarily members of those Committees. Open discussion of committee reports is encouraged.

Scope of S2: Standards, specifications, methods of measurement and test, and terminology in fields of mechanical vibration and shock, and condition monitoring and diagnostics of machines, but excluding those aspects which pertain to biological safety, tolerance and comfort.

Session 3pAA

Architectural Acoustics and Noise: Cruise Ship Acoustics (High Seas Acoustical Adventures)

Russell A. Cooper, Cochair
Jaffe Holden Acoustics, 114A Washington Street, Norwalk, Connecticut 06854

Neil A. Shaw, Cochair
Menlo Scientific Acoustics, P.O. Box 1610, Topanaga, California 90290-1610

Chair's Introduction—1:30

Invited Papers

1:35

3pAA1. Audio system design for cruise ships. Alan Edwards (Nautilus Entertainment Design, Inc., 470 Nautilus St., Ste. 301, La Jolla, CA 92037)

The design of audio systems for cruise ships introduce several acoustic challenges. The two leading challenges to overcome are ambient noise level and specular reflections. Ambient noise levels can reach 55 dBA and in many cases, higher. A majority of these noises are caused by engine vibrations, and HVAC which tends to be the leading cause of high noise floor levels. Audio systems are designed to compensate for such high ambient noise levels by distributing sound levels selectively throughout the facility. Another challenge to audio system design is specular reflections. This reflection is a delayed audible signal bounced off a surface at equal to or slightly less than the level of the direct sound with time as the variable. The problems caused by these reflections can vary from tonal coloration to echoes. Many surface materials not found in traditional showrooms can have very low absorption coefficients around the 1000 Hz–4000 Hz margin. Such surface materials include heavy plated glass windows, smooth tile flooring, etc. Interaction with the Interior Architect is crucial when designing the audio system. Many of these common problems can be minimized by room modeling through computer simulators for acoustics in the design phase.

1:55

3pAA2. Sound insulation performance in shipboard. Hyun-Ju Kang, Sang-Ryul Kim, and Jae-Seung Kim (Acoust. Res. Lab., Korea Inst. of Machinery and Mater., P.O. Box 101, Yusung, Taejon 305-600, Korea)

This work deals with correlation between the field sound transmission loss (FSTL) in cabins and the sound transmission loss (STL) in reverberation chambers. To this end, experimental investigations are carried out on various ships such as a passenger, a plant ship, and a container ship. Comparisons of FSTLs and STLs indicate that a passenger ship shows the best performance in sound insulation over the other ships, with equal or greater than STL. This is because the factory made unit cabins in the passenger ship are free from sound leakage, while the common partitions between cabins in other ships cannot be free from flanking transmissions. It is found that FSTL can be degraded up to 15 dB in comparison with STL, mainly because of sound leaks. Numerical predictions are carried out in order to identify each sound leak between structural members such as wall and ceiling. Agreement between experiments and predictions is reasonably good with the discrepancy up to 3 dB in general, which validates the proposed scheme for evaluating FSTL.

Contributed Papers

2:15

3pAA3. The acoustic design for theaters on cruise ships: The differences between land and sea. Russell Cooper (Jaffe Holden Acoust., Inc., 114A Washington St., Norwalk, CT 06854, rcooper@jhacoustics.com)

Theater design for cruise ships presents many challenges that are different than building designs on land. The acoustic challenges include a limited palette of material choices based on weight, space, and fire restrictions, room adjacency issues, and coordination with shipyard construction methods. We will discuss the program uses of the spaces and the acoustic criteria, architectural treatments, room shaping and volume, sound isolation utilizing sandwich and floating floor constructions, and HVAC noise control without the use of fiberglass. We will specifically focus on the design for Royal Caribbean Cruise Lines Voyager, Millennium, and Radiance class ships. Information will be presented that could be transferable to land based designs to solve acoustic problems that involve similar weight, space, and fire restrictions.

2:30

3pAA4. Electro-acoustic design: The differences between land and sea. Howard Rose (Jaffe Holden Acoust., Inc., 114A Washington St., Norwalk, CT 06854, hrose@jhacoustics.com)

For the past 5 years Jaffe Holden Acoustics has been consulting with Royal Caribbean Cruise Lines on the design of their theaters on board their latest three classes of cruise ships; Voyager Class (currently the worlds largest class of cruise ship), Millennium Class, and Radiance Class. Designing performance audio systems for cruise ships presents a variety of design and installation challenges. Some of the challenges we will discuss include: space restrictions, cabling practices, mounting methods, power requirements and restrictions, trade coordination, client politics, and budget. Given this multitude of challenges, we will discuss the success of these designs. We will present case studies for a few ships, some commissioned and some still under construction.

2:45

3pAA5. Shipboard noise prediction algorithms. Raymond W. Fischer (Noise Control Eng., Inc., 799 Middlesex Turnpike, Billerica, MA 01821)

For new ship designs to have acceptable noise environments, it is imperative that effective acoustical prediction tools be used. Accurate prediction tools can be used to optimize the selection of treatments and aid in trade-off studies. In this manner the typical adverse impact of noise control treatments on space, weight, cost, and operation of the vessel can be minimized. This paper describes the algorithms used in a hybrid Statistical Energy Analysis (SEA) airborne noise prediction program specifically geared for marine applications. The program combines the elements of architectural acoustic, energy analysis and empirical data collected over many ship projects. It accounts for the unique features involved with marine structures and is applicable to most types of large and small vessels mono-hulls, catamarans, surface effect ships (SESs) or Small Water Area Twin Hulls (SWATHs). Not only does the program predict the noise levels it can be used to determine the primary transmission paths, vibration levels, and radiation efficiency of any structural component, and composite panel transmission loss. The program has been employed with great success on over a dozen projects. Comparison of predicted and measured noise levels will be presented.

WEDNESDAY AFTERNOON, 5 DECEMBER 2001 ROOM 305, 1:00 TO 2:45 P.M.

Session 3pAB

Animal Bioacoustics, Underwater Acoustics, Acoustical Oceanography and Noise: Acoustics and Marine Mammals I: Perspective on Noise and Marine Mammals

Joseph E. Blue, Cochair
3313 Northglen Drive, Orlando, Florida 32806

Edmund R. Gerstein, Cochair
Leviathan Legacy, Inc., 1318 Southwest 14th Street, Boca Raton, Florida 33486

Chair's Introduction—1:00

Invited Papers

1:05

3pAB1. The regulatory agency's perspective on noise and marine mammals. Roger L. Gentry (NOAA, Natl. Marine Fisheries Service, Silver Spring, MD 20910)

The community is a little closer to resolving this issue now than it was at the 1999 Columbus symposium. The same laws, guidelines, agencies, researchers, and sources of money apply now as before, only now results are limited more by a shortage of researchers than of money. A few important advances have been made in research. The use of these results by regulators to make decisions is being opposed by a new mistrust of all science conducted by the government. Several interwoven factors, new since March 2000, are affecting the way the marine mammal/noise issue is being managed. These include the Bahamian beaked whale stranding, use of the general public as a weapon to force decisions on complex acoustic issues, and increasing interest in acoustics by federal legislators. This paper will discuss these new factors, and will guess at changes that may occur in the near future.

1:25

3pAB2. Regulating ocean noise: Limits and possibilities of existing legal frameworks. Joel R. Reynolds and Michael D. Jasny (Natural Resources Defense Council, 6310 San Vicente Blvd., Ste. 250, Los Angeles, CA 90048)

This paper considers the framework that existing federal laws provide for regulating ocean noise and evaluates amendments that have been proposed with the intention of improving that framework. Particular attention is paid to the U.S. Marine Mammal Protection Act (MMPA), which establishes a moratorium on the ''harassment'' of marine mammals, and to changes in its language that have recently been recommended by scientific committees or advanced by administrative agencies through *ad hoc* rule making. The adverse consequences that some of these changes would have, as for example by hampering the agency's ability to enforce the law, by restricting public notice, and by undermining the process of judicial review—not only with regard to ocean noise pollution but to other environmentally harmful activities that come under the purview of the MMPA—have not adequately been considered. Finally, this paper discusses alternatives to the existing regime, including steps that other governments have taken (such as setting noise limits in marine protected areas); and opportunities available for addressing ocean noise pollution under international conventions and protocols.

1:45

3pAB3. Science and the management of underwater noise: Information gaps and polluter power. Hal Whitehead and Linda Weilgart (Dept. of Biol., Dalhousie Univ., Halifax, NS B3H 4J1, Canada, hwhitehe@is.dal.ca)

To regulate underwater noise rationally and efficiently, we need to know its impact on population and community biology. This link can rarely be studied directly because of logistics and the spatial and temporal scales of ecological processes. There are two principal approaches: correlational studies of noise levels with population events or measures; and experiments in which the response

variables (usually short-term behavioral measures) are poor proxies for the population and ecosystem parameters about which we are concerned. Experimental studies also have costs. These may include the introduction of additional noise, delay of substantive regulation, and, when polluters are major funders, perceived gagging or biasing of knowledgeable scientists. This is a particular problem with underwater noise because the U.S. Navy (a major noise polluter) and allied organizations fund so much acoustic research. Consequently (a) managers must recognize that underwater noise is dangerous but its most important consequences cannot currently be determined; (b) following the precautionary approach, noise levels should be reduced, sources distanced from marine life, and new noises avoided; (c) correlational studies are generally preferred to experimental ones; (d) major noise polluters should not directly fund the research, instead providing fees to independent bodies which commission research and recommend regulations.

2:05

3pAB4. Interpreting research results: Government regulation of anthropogenic noise sources. Naomi A. Rose (The Humane Society of the U.S., 2100 L St. NW, Washington, DC 20037)

The Precautionary Principle, which advocates taking anticipatory management action in the absence of complete proof of harm, is embedded in several U.S. environmental laws, including the Marine Mammal Protection Act and the Endangered Species Act. However, in practice it is often ignored. Recent U.S. regulatory actions addressing the following anthropogenic noise sources are evaluated as cases in point: the North Pacific Acoustic Laboratory (NPAL), the U.S.S. Winston S. Churchill ship shock trials, and the Surveillance Towed Array Sensor System (SURTASS) Low Frequency Active (LFA) sonar. The impact of anthropogenic noise on marine mammals is poorly understood. Therefore, establishing adequate mitigation measures for noise sources is difficult. In an effort to increase knowledge and build toward certainty regarding the impacts of anthropogenic noise on the marine environment, recent research on low-frequency noise has focused on behavioral responses of large whales. These studies address limited hypotheses, but their results have been broadly interpreted to justify regulatory decisions. The pursuit of carefully designed research on the impacts of noise on the marine environment and the refinement of noise source mitigation is important, but interpreting research results is subject to political and economic pressures, casting doubt on the research conservation value.

2:25

3pAB5. A journalist's perspective on noise and marine mammals. Dick Russell (8118 Hollywood Blvd., Los Angeles, CA 90069, dickrusl@aol.com)

In the course of several years researching and writing *Eye of the Whale* (Simon & Schuster, 2001, 688 pages), the author interviewed a number of marine scientists about this subject and described at some length scientific findings over the past several decades. These pertained to California gray whales, their vocalizations, and the impact of various acoustical ''playbacks'' upon this species at different locations. In particular, the author described studies of Navy LFA sonar in California and of seismic oil exploration off Russia's Sakhalin Island. Scientific opinion appeared divided over whether LFA sonar, especially, would pose a substantial threat to marine mammals. However, as regarding both LFA sonar and seismic testing, gray whales were observed to avoid certain regions they would otherwise normally inhabit. Subsequent interviews have confirmed the possibility of a severe threat to marine mammal habitat. During this process, much has also been learned about the ways and means to improve communication between journalists and scientists.

WEDNESDAY AFTERNOON, 5 DECEMBER 2001 ROOMS 203/204, 1:15 TO 3:15 P.M.

Session 3pED

Education in Acoustics: "Take Fives:" Sharing Ideas for Teaching Acoustics

Uwe J. Hansen, Chair
Department of Physics, Indiana State University, Terre Haute, Indiana 47809

Chair's Introduction—1:15

Do you have a novel demonstration, a new laboratory experiment, a favorite video, a recorded sound example, or a new idea for teaching acoustics which you are willing to share with your colleagues? At this session a sign-up board will be provided for scheduling presentations. No abstracts are printed. Presenters are encouraged to have handouts to distribute. Multiple presentations are acceptable (not consecutively). Presentations are limited to 5 minutes. Keep them short! Keep them fun!

Session 3pID

Interdisciplinary: Hot Topics in Acoustics

Ralph T. Muehleisen, Chair
Civil Environmental and Architectural Engineering, University of Colorado, Box 428, Boulder, Colorado 80309

Chair's Introduction—2:00

Invited Papers

2:05

3pID1. Hot topics in speech communication. Peter F. Assmann and Philipos C. Loizou (School of Human Development, Univ. of Texas, Dallas, Box 830688 GR 41, Richardson, TX 75083-0688)

Speech communication is a robust and flexible process. For listeners with normal hearing, speech remains intelligible when a substantial proportion of the spectrum is eliminated by filtering, when portions of the wave form are replaced with silence, and in the presence of interfering sounds, including sounds made by competing talkers. The resilience of speech depends on several interacting systems which exploit the redundancy of speech in different ways. New results are reported on the perceptual processing of talker variability, monaural and binaural mechanisms of sound separation, audio-visual integration in lip reading, and perceptual compensation for channel distortion. For listeners with hearing impairment, new research in the field of cochlear implants has focused on the development of new electrode arrays, the efficacy of bilateral implants and the development of sophisticated signal processing techniques to combat noise. The new electrode arrays are designed to limit electrode interaction and thus increase the number of perceptual channels of information. The benefit of bilateral implants is investigated in terms of sound localization and speech perception in noise. New results are reported on bilateral cochlear implants, speech-processing strategies for noisy environments and acoustic simulations of cochlear implants which aim to quantify the factors contributing to speech intelligibility.

2:25

3pID2. Hot topics in physical acoustics. Robert M. Keolian (Grad. Prog. in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804-0030, bonzo@sabine.acs.psu.edu)

With the adoption of the 1996 Comprehensive Nuclear-Test-Ban Treaty, a global International Monitoring System of 321 monitoring stations is being deployed around the world. Three of the four types of sensors used by these stations are acoustical in nature. Seismic, hydroacoustic, infrasonic, as well as radionuclide sensors are used to detect banned underground, underwater, and atmospheric nuclear tests. The infrasonic sensors are arrays 1 to 3 kilometers across to lower their sensitivity to noise from the wind. They can detect nuclear blasts thousands of kilometers away, but they also detect some background events that are interesting too: meteorites, explosive volcanoes, supersonic aircraft, and reentering space debris.

2:45

3pID3. Hot topics in animal bioacoustics. Mardi C. Hastings (Biomed. Eng. Ctr., Ohio State Univ., 1080 Carmack Rd., Columbus, OH 43210, hastings.6@osu.edu)

Recent discoveries in the field of animal bioacoustics include everything from active auditory mechanics in insects to the behavior of large marine mammals in response to manmade sounds. Using laser Doppler vibrometry, Göpfert and Robert found that active auditory mechanics in male mosquitoes provides tuning that allows them to detect flying females. Their exciting discovery revealed new similarities between insect and vertebrate hearing. Many researchers have reported the behavior of fishes and marine mammals in response to underwater manmade sounds that is most likely controlled by other factors than the sound itself. Wardle and his colleagues found that repeated firing of seismic air guns had little effect on the day-to-day behavior of reef fishes off the coast of Scotland unless the explosions were visible. Madsen and Møhl also reported no change in behavior or click rate of sperm whales when detonators were discharged off the coast of Norway. Thus marine animals may not swim away from anthropogenic sounds even though they could be harmful. Gerstein determined masked thresholds and critical ratios for manatees that may prevent them from detecting boat noise in their natural habitat. His findings will lead to new solutions enabling manatees to detect boats and avoid collisions.

Session 3pNS

Noise: Ray-Tracing Software for Predicting Community Noise from Industrial Facilities II: Panel Discussion

Marlund E. Hale, Chair
Parsons Engineering Science, Inc., A2-04-A, 663 Bristol Avenue, Simi Valley, California 93065-5402

Chair's Introduction—1:15

Ray-tracing and similar image source software is becoming increasingly popular for predicting industrial noise in the community. Ray tracing can readily handle the effects of screening and reflections by barriers, buildings and other equipment, which are usually difficult using other prediction methodologies. Available software has great power and precision, which is gained at the expense of time and complexity. This power can quickly get users into problems, often without their knowing it. The current software has many limitations and problems, particularly when including excess attenuations and wind effects. The special session is intended to provide insight into the available software, as well as some practical guidance on using it. Such an understanding is a prerequisite to effectively using the software. The special session will be especially useful for noise control engineers who are currently using community noise prediction software or who are considering using it. This panel discussion will present an opportunity for attendees to both make comments regarding their experiences with predicting community noise and to ask questions of one another. An opportunity will be provided for the invited speakers and the attendees to answer questions. This panel discussion will provide an excellent opportunity to share information, and learn from others actively involved in using this software.

Session 3pUW

Underwater Acoustics: Matched Field Techniques and Applications

J. Mark Stevenson, Chair
Space and Naval Warfare Systems Center, 49575 Gate Road, San Diego, California 91252

Contributed Papers

1:00

3pUW1. Environmentally robust nulling applied to rank-N stochastic matched field processing. Peter M. Daly (SAIC Ocean Systems Div., peter.m.daly@saic.com) and Arthur B. Baggeroer (MIT)

Passive source localization in shallow water is complicated by both the uncertain propagation environment and the presence of strong interferers. Traditional spatial nulling techniques fail to accommodate environmental variability, while principle component methods neglect energy that has spread beyond a rank-1 signal characterization. One can accommodate environmental variability by treating the parameters of a propagation environment as random quantities. This yields a random acoustic pressure field at a receiver, which can be approximated by a Gaussian random vector, represented by a spatial covariance matrix. To mitigate these nulling problems, one can characterize the interfering signal in this manner. With such a model, a whitening filter can be calculated and applied to the received signal. This provides a spatial null that removes the multiple-component energy contributed by the interferer in an uncertain propagation environment. Rank-N localization techniques can then be applied, robustly estimating a target source location in the presence of environmental uncertainty. [Work supported by the DARPA Advanced Technology Office.]

1:15

3pUW2. Quantitative ambiguity analysis for matched-field parameter estimation. Wen Xu and Arthur B. Baggeroer (MIT, 77 Massachusetts Ave., Rm. 5-204, Cambridge, MA 02139)

Matched-field methods concern the estimation of source location and/or ocean environmental parameters by matching the measured signal field with the modeled signal field. The resulted ambiguity output is often characterized by a multimodal structure since the typical signal field is a highly nonlinear function of the embedded parameters. In the high signal-to-noise ratio (SNR) region, the peak of the true parameter protrudes prominently above the noise and can be located accurately; below some SNR, the true peak often falls below the noise level and is obscured by other ambiguous peaks, leading to a larger estimation error and the well-known threshold phenomenon. Therefore, issues on the output ambiguity structure are very important for the development of any matched-field algorithm operated in low SNR scenarios (e.g., passive source localization). To understand some fundamental ambiguity behaviors, a quantitative approach for error analysis is developed in the context of the maximum likelihood estimate. The derived probabilistic error at each scanning parameter point is determined by the signal ambiguity level, the distance to the true parameter point, and the SNR. Accordingly, the threshold SNR can be located by comparing the relative size of the main lobe and side lobe error contributions.

1:30

3pUW3. Cramer-Rao bounds for source localization and matched field tomography in a partially stratified ocean. Arthur B. Baggeroer (MIT, Cambridge, MA 02139)

The performance of Matched Field Processors (MFP) for either source localization or tomography using Cramer–Rao bounds often gives overly optimistic results. This is often because the analyses to date have used fully coherent models for the acoustic propagation. This implies that the modes or rays remain sychronized, or ''phase locked.'' This is a built in assumption for most mode/ray codes and parabolic equation models. Since the phase interaction provides much of the information for MFP, any environment that leads to randomization of the mode/ray coupling cannot use a single degree of freedom replica for optimal processing. Models that address the randomness are often described as ''partially saturated'' propagation. Moreover, the closed form formulas for MFP performance analysis using Cramer–Rao bounds are not applicable for such models. Since partially saturated acoustic propagation is often a more realistic model, it is important to determine the performance of MFP in this regime. Here, we derive closed form formulas for computing Cramer bounds and provide some examples of the degradation due to partial saturation. [Work supported by ONR Code 321: Undersea Signal Processing.]

1:45

3pUW4. Matched field processing and tracking with sparsely populated line arrays. Ahmad T. Abawi and J. Mark Stevenson (SPAWAR Systems Ctr., San Diego, CA)

Data from the Shallow Water Cell Experiments SWellEx-96, which was performed in roughly 200 m of water, are used to assess the detection and localization performance of sparsely populated line arrays. The acoustic source in this experiment transmitted 13 narrow-band tones between 50–400 Hz. Data from this source was recorded on a 64-phone vertical line array with a vertical aperture of 118 m and a 240-meter long, 32-phone horizontal line array. Matched field processing and tracking using these data are performed and the detection performance of these arrays as a function of source level and the number of phones is investigated.

2:00

3pUW5. A test of autonomous matched-field tracking with a sparse seafloor array. J. Mark Stevenson, Homer Bucker, Susan G. Briest, Mike Klausen, and Vincent K. McDonald (Space and Naval Warfare Systems Ctr., San Diego, CA 92152)

A recent experiment demonstrated the feasibility of using input from a sparse seafloor array of hydrophones [Stevenson and Bucker, J. Acoust. Soc. Am. **108**, 2645 (2000)] into a matched-field tracking algorithm [Bucker, J. Acoust. Soc. Am. **96**, 3809–3811 (1994)] to return tracks of submerged sound sources. In previous work, the algorithm was tested using recorded data. This paper describes a near real-time implementation of this processing. The tracking algorithm operates in near real-time in a DSP (digital signal processor) running autonomously inside the unit. The procedure is to select N Fourier frequency bins based on S/N ratios after CFAR normalization, and then to determine the best tracks using M-of-N processing. Observations over 3–6 min tracks are matched against the calculated sound field stored in a look-up table [Bucker and Baxley, J. Acoust. Soc. Am. **106**, 3226–3230 (1999)]. We compare near real-time results obtained autonomously with those obtained in post-experiment analysis. [Work supported by ONR, Code 321-SS.]

2:15

3pUW6. Source motion compensation using waveguide invariant theory. Lisa M. Zurk (MIT, Lincoln Lab., 244 Wood St., Lexington, MA 02420)

Passive detection and localization of quiet, submerged sources moving in shallow water in the presence of shipping noise is a critical problem for the ASW community. Past work has demonstrated that adaptive processing can be applied to detect and localize stationary sources while mitigating loud interferers. However, source motion can cause significant signal gain degradation, loss of localization accuracy, and incomplete interference rejection. The severity of the motion loss is channel dependent, and a function of the amount of motion during the observation interval. Motion compensation techniques have been demonstrated in past work with data from the Santa Barbara Channel Experiment. These algorithms utilized a normal mode propagation model and a velocity hypothesis to successfully compensate for source motion. However, the technique requires good environmental knowledge and can be computationally expensive if a large hypothesis space must be searched. In this paper, the waveguide invariance principle is used to combine the time-frequency snapshots acquired from a broadband source to compensate for motion. The technique does not require knowledge of the underwater environment for the compensation and the search across the velocity space can be structured so it is computationally efficient. Results from this technique will be presented and discussed.

2:30

3pUW7. Shallow water localization of targets in range rate by use of sequential importance sampling. Shawn Kraut and Jeffrey Krolik (Dept. of ECE, Duke Univ., Box 90291, Durham, NC 27708-0291, jk@ee.duke.edu)

We examine the problem of localization of targets in shallow water environments. In contrast to standard matched field processing (MFP) approaches, we propose a method based on range-rate, rather than range. We call this approach moving target range-rate estimation (MOTRE). It is more robust with respect to environmental uncertainty. This is because range-rate localization depends on the product of the horizontal wave numbers of the channel modes times the differential distance the target travels in a sampling interval, rather than the baseline distance to the sensor array. Our technique assumes some knowledge of target dynamics (constant velocity) and the physics of the channel modes to hypothesize candidate sequences of mode coefficients. A likelihood surface in range-rate and depth is computed using a nonlinear state-space model, related to a Kalman filter, to constrain the evolution of the unknown mode coefficients. We update the likelihood as new data is received by using a method of sequential importance sampling (SIS), a topic of interest in recent statistics literature. Our results indicate that MOTRE outperforms conventional MFP in its ability to produce easily identifiable, unambiguous peaks in the likelihood surface, using a vertical line array in a shallow-water channel.

2:45

3pUW8. Automated model-based localization of marine mammals near Hawaii. Christopher O. Tiemann, Michael B. Porter (Sci. Applications Intl. Corp., 1299 Prospect St., Ste. 303, La Jolla, CA 92037), and L. Neil Frazer (Dept. of Geology and Geophys., Univ. of Hawaii at Manoa, Honolulu, HI 96822)

Acoustic data from the Pacific Missile Range Facility (PMRF) hydrophone network near Hawaii is being used to develop a real-time automated alert and tracking algorithm. The sources of interest are marine mammals, humpback whales in particular, which are regularly seen in the vicinity of the PMRF array. The algorithm under development uses acoustic data from six hydrophones plus an acoustic propagation model to construct an ambiguity surface identifying the most probable whale location in a horizontal plane around the array. It has the further advantage that it can be implemented in real-time and without human interaction, making it suitable for automated alert applications.

3:00

3pUW9. The mudslide effect. Frederick D. Tappert (Div. of Appl. Marine Phys., Univ. of Miami, RSMAS, 4600 Rickenbacker Cswy., Miami, FL 33149), John L. Spiesberger (Dept. of Earth and Environ. Sci., Univ. of Pennsylvania, 240 S. 33rd St., Philadelphia, PA 19104-6316), and Michael A. Wolfson (Appl. Phys. Lab., College of Ocean and Fishery Sci., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698)

A novel range-dependent propagation effect occurs when a source is placed on the seafloor in shallow water with a downward refracting sound speed profile, and sound waves propagate down a slope into deep water. Small grazing-angle sound waves slide along the bottom downward and outward from the source until they reach the depth of the sound channel axis in deep water. They then detach from the sloping bottom and continue to propagate outward near the sound channel axis. This ''mudslide'' effect is one of a few robust and predictable acoustic propagation effects that occurs in range-dependent ocean environments. Consequently, a bottom mounted source in shallow water can inject a significant amount of energy into the axis of the deep ocean sound channel that propagates to very long ranges. Numerical simulations with a full-wave range-dependent acoustic model show that the Kaneohe experiment had the appropriate source, bathymetry, and sound speed profiles that allowed this effect to operate efficiently. This supports the interpretation that some of the near-axial acoustic signals, received near the coast of California from the Kaneohe source on Oahu, Hawaii, were injected into the sound channel of the deep Pacific Ocean by this mechanism. [Work supported by ONR.]

WEDNESDAY AFTERNOON, 5 DECEMBER 2001 BALLROOM A, 3:30 TO 4:30 P.M.

Plenary Session and Awards Ceremony

William M. Hartmann, President

Acoustical Society of America

Presentation of Certificates to New Fellows

Shira L. Broschat

René Caussé

Pierre L. Divenyi

Mathias Fink

Anthony G. Galaitsis

Paul C. Hines

Anatoliy N. Ivakin

Jerry G. Lilly

Chaslav V. Pavlovic

Allan G. Piersol

Ahmet Selamet

Michael Taroudakis

William A. Watkins

Presentation of Awards

Science Writing Award in Acoustics for Journalists to Graham Lawton

Science Writing Award for Professionals in Acoustics to Colin Gough

Science Writing Award for Professionals in Acoustics to Thomas D. Rossing

Medwin Prize in Acoustical Oceanography to Timothy G. Leighton

Silver Medal in Engineering Acoustics to Ilene J. Busch-Vishniac

Session 4aAA

Architectural Acoustics and Speech Communication: Speech Intelligibility and the Metrics Used for its Evaluation I

Kenneth P. Roy, Cochair
Innovation Center, Armstrong World Industries, 2500 Columbia Avenue, Lancaster, Pennsylvania 17603

Peggy B. Nelson, Cochair
Communication Disorders, University of Minnesota, 115 Shevlin Hall, 164 Pillsbury Drive, SE, Minneapolis, Minnesota 55455

Chair's Introduction—8:30

Invited Papers

8:35

4aAA1. Acoustic considerations for speech intelligibility by special populations. Peggy Nelson (Dept. of Commun. Disord., Univ. of Minnesota, 164 Pillsbury Dr. SE, Minneapolis, MN 55455)

Increasingly the design of public spaces is complicated by the communication needs of multiple groups of users. Schools, places of worship, and community centers are used for meetings of children and adults, with and without hearing loss, who are native speakers of English or other languages. In this presentation, data concerning the acoustic needs of children, speakers of other languages, and listeners with hearing loss will be summarized. Data suggest that typically developing children require more favorable acoustic conditions than do adults. Listeners with hearing loss and those listening in a second language need even more favorable acoustics for speech intelligibility. Desired acoustic guidelines for public spaces will be described. Emphasis will be placed on the need for quieting noise and reducing reverberation, rather than amplifying sound.

9:00

4aAA2. Assessment of the classroom as a communication channel. Sigfrid D. Soli and Jean Sullivan (Dept. of Human Commun. Sci. and Devices, House Ear Inst., Los Angeles, CA 90057)

The classroom is an environment in which spoken language communication facilitates and enables students in the learning of essential academic, social, and cultural skills. As such, the classroom defines a communication channel for learning. Our research has focused on the special requirements placed on this communication channel by the student listener. Our strategy is to quantify the effects of age, developmental status, hearing status, and English proficiency on speech intelligibility in noise and to find the signal/noise ratios that enable all students to obtain high intelligibility. Using these signal/noise ratios, acceptable noise levels and reverberation times for the classroom communication channel are estimated. Our studies of young children for whom English is either the primary language (EPL) or the secondary language (ESL) will be reported. The method of estimating the special requirements these children place on the classroom communication channel will also be described. Acoustic measurements from six elementary school classrooms in which ESL and EPL children are taught will be presented, and these classrooms will be evaluated in relation to the communication channel requirements.

9:25

4aAA3. Computer model studies of RASTI values in classrooms. Gary W. Siebein, Bumjun Kim, Hyun Paek, and Mark LoRang (Architecture Technol. Res. Ctr., 231 ARCH, Univ. of Florida, P.O. Box 115702, Gainesville, FL 32611-5702)

Computer models of classrooms were used to study the effects of room shape, furniture arrangement and interior finish materials at various background noise levels. The background noise levels used in the computer models were obtained in field measurements of actual classrooms and varied from approximately Noise Criterion (NC) 25 to NC 65. A sound source was configured to approximate the level, frequency and directionality of the normal speaking voice of a teacher in the room. RASTI, reverberation time, loudness, and early energy ratios were estimated for multiple seating locations in each room for each background noise level. RASTI was related to background noise level (NC rating), distance and room finish materials with a second order equation with a model r^2 of 0.96. The background noise level had the primary effect on RASTI values in the rooms. Once a background noise level of NC 35 or less was achieved, the amount and distribution of absorbent materials in the room became important factors. Reflective materials in parts of the ceiling of the rooms contributed to increased loudness and higher RASTI values at most seating locations.

9:50

4aAA4. The effect of wall angles near a speech source on hearing comfort. Abigail Stefaniw and Yasushi Shimizu (Architecture Dept., Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, stefaa@alum.rpi.edu)

This study will involve a controlled experiment to determine the correlation between hearing comfort and lateral energy fraction (LEF). (Hearing comfort has also been described as ease of hearing, and is a measure of speech intelligibility.) LEF is dependent on the geometry of surfaces near the source. Therefore, in order to study the effect of the angle of reflective walls near a speech source on hearing comfort, rooms with identical reverberation time, noise level, and volume are modeled. The angle between walls near the speaker is the only physical variable. The RT in every room will be 0.6 s and the background noise will be exactly NC 25. The angle between reflective walls on either side of the speaker will be changed in 5° increments. The speaker is always located 3 ft. in front of the varying corner, and the listener is always located 15 ft. from source. This ensures that the signal to noise ratio can also be held constant. The data from this modeling experiment should explain the relationship between hearing comfort and room geometry, as measured by the lateral energy fraction.

10:15–10:25 Break

10:25

4aAA5. A completed study of the architectural repair of the acoustics in Koubek auditorium at the Catholic University of America. Donna A. Ellis (The Catholic Univ. of America, 521 Manor Rd., Severna Park, MD 21146, dellisdona@aol.com)

Koubek Auditorium is used for architecture classes and guest lectures in the School of Architecture and Planning at the Catholic University of America in Washington, DC. Inadequate acoustics in this space result in ineffective use of the space as well as a poor teaching environment. The main problem areas are excessive background levels and reverberation times. Using a 1/3-octave band, background levels range from 47.6 dB–56.5 dB, reverberation times range from 0.9–1.6 s, and the noise criteria (NC) is anywhere from NC40 to as high as NC60, depending on where one is seated. The intent of this paper is to discuss the analysis and implementation of the acoustical corrections, as well as the end result of the changes and how those results improve the learning environment in the space.

10:50

4aAA6. Effects of background-noise on speech listening difficulty in reverberant sound fields. Masayuki Morimoto, Masaaki Kobayashi, Atsushi Hakamada (Environ. Acoust. Lab., Kobe Univ., Nada, Kobe 657-8501, Japan), and Hiroshi Sato (Grad. School of Eng., Tohoku Univ., Aoba, Sendai 980-8579, Japan)

The previous work [M. Morimoto *et al.*, J. Acoust. Soc. Am. **108**, 2633 (2000)] demonstrated that the listening difficulty can evaluate the quality of speech transmission more exactly and sensitively than intelligibility. In the present work two listening tests were performed to investigate the effects of background noise on the listening difficulty in reverberant sound fields. In the tests, both the listening difficulty and the intelligibility were assessed using word lists based on word familiarity. In the first test, the background noise level was changed, while the speech level was constant at 55 dBA. In the second test, both the speech and background noise levels were changed in the same manner, keeping the signal to noise ratio constant. The results of both tests supported the previous results. Furthermore, the results of the first test indicated that the difficulty did not always increase when background noise levels exceeded the absolute threshold of the noise. Namely, background noise lower than about 30 dBA does not increase the difficulty. The results of the second test demonstrated that the optimum signal to noise ratio depended on background noise level. Overamplifying speech to keep the signal to noise ratio constant increases the difficulty, though the intelligibility does not decrease.

11:15

4aAA7. Speech listening difficulty and word intelligibility of aged listeners in reverberant sound fields. Hiroshi Sato, Hayato Sato, Hiroshi Yoshino (Grad. School of Eng., Tohoku Univ., Aoba, Sendai 980-8579, Japan), and Masayuki Morimoto (Environ. Acoust. Lab., Kobe Univ., Nada, Kobe 657-8501, Japan)

The experiment was performed to present the effect of hearing loss on listening difficulty and word intelligibility with aged subjects in same sound fields of the previous work [M. Morimoto *et al.*, J. Acoust. Soc. Am. **108**, 2633 (2000)], and the result of this experiment is compared with this previous work. In the test, both the listening difficulty and the intelligibility were assessed using word lists based on word familiarity. The results demonstrate that (1) The relation between listening difficulties of young subjects and aged subjects is highly correlated. The range of scores of aged subjects is half of that of young subjects. (2) This word intelligibility score of young subjects is higher than that of aged subjects by 25% in all conditions. This result is the same as the previous work [H. Sato *et al.*, J. Acoust. Soc. Am. **108**, 2634 (2000)]. (3) The relation between listening difficulty and word intelligibility shows the same tendency for the both aged and young subjects. (4) Hearing level has a good correlation with listening difficulty and word intelligibility scores.

Session 4aAB

Animal Bioacoustics, Underwater Acoustics, Acoustical Oceanography and Noise: Acoustics and Marine Mammals II: Acoustic Detection and Underwater Propagation I

Darlene R. Ketten, Chair
Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Chair's Introduction—8:00

Invited Papers

8:05

4aAB1. Review of marine mammal temporary threshold shift (TTS) measurements and their application to damage-risk criteria. James J. Finneran (Sci. Applications Intl. Corp., Maritime Services Div., 3990 Old Town Ave., Ste. 105A, San Diego, CA 92110), Donald A. Carder, and Sam H. Ridgway (Space and Naval Warfare Systems Ctr., San Diego, CA 92152-5000)

Intense anthropogenic underwater sound may adversely effect the hearing and behavior of many marine mammals. Exposure to intense sound may produce an elevated hearing threshold, also known as a threshold shift (TS). If the threshold returns to the pre-exposure level after a period of time, the TS is known as a temporary threshold shift (TTS); if the threshold does not return to the pre-exposure level, the TS is called a permanent threshold shift (PTS). PTS and TTS data were used to establish noise exposure limits in humans and TTS data are often used to help predict the effects of anthropogenic noise on marine mammals in the wild. Previous studies of TTS in marine mammals conducted at SSC San Diego have examined the effects of exposure to single 1 s pure tones and underwater impulsive waveforms. This paper reviews the TTS studies conducted at SSC San Diego over the last 5 years and summarizes the resulting data. The application of these data to the development of damage-risk criteria for marine mammals will also be discussed.

8:25

4aAB2. Aging, injury, disease, and noise in marine mammal ears. D. R. Ketten, J. O'Malley (Woods Hole Oceanogr. Inst., Woods Hole, MA and Harvard Med. School, Boston, MA, dketten@whoi.edu), P. W. B. Moore, S. Ridgway (SPAWAR, San Diego, CA 92152), and C. Merigo (New England Aquarium, Central Wharf, Boston, MA 02110)

In humans, hearing is lost with age, trauma, and disease. It is likely marine mammals sustain similar damage, but currently we have little direct data on their incidence of hearing loss. Marine mammals represent an interesting hearing paradox. Because their ears are fundamentally the same as human ears, we expect they have similar loss mechanisms, but they evolved in a high noise environment and have adaptations that deter pressure-induced trauma. These adaptations may equate to ''tougher'' ears. This paper summarizes evidence from computerized tomography and histologic exams for grades and etiology of hearing loss in captive and wild marine mammals. The data indicate dolphins have hypertrophic inner ears elements (i.e., stria vascularis, spiral ligament, basement membrane) that imply greater robustness and faster recovery times than pinniped ears. However, older dolphins and seals alike develop degenerative pathologies, including neural, hair cell, support cell, and bony loss that parallel presbycusic changes in older humans and are consistent with substantial hearing loss. The results suggest whales and dolphins may be less subject to temporary threshold shifts than pinnipeds, but both are equally liable to hearing loss from severe trauma or age. [Work supported by ONR and NOAA Fisheries.]

8:45

4aAB3. Measuring recovery from temporary threshold shifts with evoked auditory potentials in the bottlenosed dolphin Tursiops truncatus. Paul E Nachtigall (Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734), Alexendre Supin (Russian Acad. of Sci., Inst. of Evolutionary Morphology, 33 Leninsky Prospekt, Moscow, Russia), Jeffrey L. Pawloski, and Whitlow W. L. Au (Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734)

Temporary threshold shifts can be short-lived in the bottlenosed dolphin and therefore difficult to measure with conventional trained behavioral psychophysical techniques. The time course of recovery from temporary threshold shifts was measured using evoked auditory potentials collected from a bottlenosed dolphin trained to wear rubber suction cups containing human EEG skin surface electrodes. During each session, following an initial measure of hearing thresholds using the evoked auditory potential procedure, the animal voluntarily positioned within a hoop 1 m underwater while 160 dB *re* 1 micropascal noise between 4 and 11 kHz was presented for 30 min. Immediately following the noise exposure, evoked thresholds were again obtained. The dolphin swam down into a second hoop located one meter in front of a calibrated hydrophone. Evoked potential thresholds were obtained 5, 10, 15, 25, 45, and 105 min following the exposure for amplitude modulated pure tones of 8, 11.2, 16, 22.5, and 32 kHz. Maximum shifts occurred 5 min following exposure and rapidly recovered. As has been observed with other animals and humans, threshold shifts depended on frequency. Shifts occurred at 8, 11.2, and 16 kHz but no shifts were detected at 22.5 and 32 kHz.

9:05

4aAB4. Pinniped hearing and anthropogenic noise. Brandon L. Southall, Ronald J. Schusterman, David Kastak, Colleen Reichmuth Kastak (Long Marine Lab., Univ. of California, Santa Cruz, 100 Shaffer Rd., Santa Cruz, CA 95060, branlee@cats.ucsc.edu), and Marla M. Holt (Long Marine Lab.)

Our behavioral studies of pinniped auditory processing, including amphibious hearing, auditory masking, and temporary threshold shift (TTS), have provided comparative data on hearing and the potential effects of anthropogenic noise for three pinniped species. Differences for one harbor seal between aerial thresholds measured using headphones in a noisy environment and those measured without headphones in a hemi-anechoic chamber are quite large. While additional data are required, the preliminary data suggest that it will be necessary to re-evaluate pinniped aerial hearing capabilities and their susceptibility to aerial anthropogenic noise impacts. Assessments of simultaneous (masking) and residual (TTS) effects of controlled noise exposure have also provided insight into the potential effects of anthopogenic noise on free-ranging pinnipeds. Our underwater masking studies indicated relatively low critical ratios for low-frequency tonal sounds, while underwater TTS studies demonstrated that auditory fatigue occurs in pinnipeds, even following exposure to moderate levels of noise. The TTS studies have also provided some means of estimating the onset of TTS and supplied data on the relative importance of fatiguing noise exposure level and duration. Results suggest that equal-energy relationships and the octave effect previously shown in humans and some terrestrial mammals may not apply for some pinnipeds.

9:25

4aAB5. It is all about SOUND science: Manatees, masking and boats. Edmund R. Gerstein, Laura A. Gerstein (Leviathan Legacy, Inc., 1318 S.W. 14th St., Boca Raton, FL 33486), Steve E. Forsythe (Navy Undersea Warfare Ctr. Div. Newport, Newport, RI 02841), and Joseph E. Blue (Leviathan Legacy, Inc., Orlando, FL 32806)

A series of controlled underwater psychoacoustic tests with captive manatees, along with comprehensive acoustic surveys of manatee habitats and boat noise propagation measurements, were conducted to investigate manatee hearing and acoustical factors that render Florida manatees vulnerable to repetitive boat collisions. Masked thresholds, critical ratios, and directional hearing of various sounds, including species-specific calls and recorded boat noise were measured against white noise and natural ambient backgrounds. Significantly high acoustic levels were required for manatees to detect approaching vessels. Slow moving boats generate relatively low power and frequency spectra compared to the louder higher frequency cavitation noise produced by faster boats. Counterintuitive to convention, the sounds of slower moving boats are more difficult for manatees to detect and they are often masked by the noise of snapping shrimp and distant faster moving boats. Furthermore, physical boundary effects inherent in shallow water attenuate and limit the propagation of dominant low-frequency sounds produced by many recreational and commercial vessels. While not adapted for hearing or localizing these lower frequencies, manatees can hear and localize an array of narrow band higher FM sounds that could be incorporated into a low-intensity directional alarm to alert them of approaching boats.

9:45–10:00 Break

10:00

4aAB6. Shallow water acoustics and marine mammals. Peter H. Rogers (G. W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30326, peter.rogers@me.gatech.edu)

Human activity is causing significant changes to the shallow water acoustic environment. A good understanding of the basics of shallow water acoustics is essential to a proper assessment of the impact of this activity on marine mammals. Misunderstanding of the relevant underwater acoustics can often lead to erroneous conclusions about the impact of such changes on marine mammal ecology. This paper will address some of the issues and misunderstandings. Shallow water regions present an acoustic environment which is typically much more demanding and complex than the deep water environment. Propagation is strongly influenced by the surface, water column, and sea bottom and will vary greatly from place to place, and time to time. Shallow water regions are highly reverberant and are often very noisy. Marine mammals rely on underwater sound for communication and detection of prey. Unlike fish, which evolved entirely in the sea, the marine mammals auditory system evolved from a terrestrial auditory system. This paper will also address the differences between shallow water acoustics and atmospheric and deep water acoustics, and how they effect the acoustic behavior of marine mammals. Specific adaptations of the dolphins echolocation system to the shallow water environment are highlighted.

10:20

4aAB7. Sound propagation in deep water. W. A. Kuperman and G. L. D'Spain (Marine Phys. Lab., Univ. of California, San Diego, La Jolla, CA 92093-0238)

The basics of deep water sound propagation as related to marine mammal issues are reviewed. The spatial distribution of sound for a variety of source situations is qualitatively and quantitatively presented. Special care will be given to the details of the units describing sound levels. Acoustic signals will also be discussed in the context of natural and manmade ambient noise.

10:40

4aAB8. A general-purpose acoustic recording tag for marine wildlife. William C. Burgess (Greeneridge Sci., Inc., 1411 Firestone Rd., Goleta, CA 93117)

Acoustic recording tags have demonstrated the potential to measure free-ranging marine animals' exposure and response to sound. To fully realize this potential, however, acoustic recording tags must be applied to a large sample of subjects. To enable the broad use of acoustic recording tag technology, a new general-purpose tag has been developed. Miniaturization to a cylinder 3 cm in diameter by 23 cm in length allows the application with a variety of species and attachment methods. Encapsulation in resin and the use of

infrared rather than electrical connectors provides durability and flexibility. The prototypes sample acoustics with 16-bit resolution at bandwidths up to 14 kHz, as well as temperature and depth with 12-bit resolution. Constant acoustic sampling at 2 kHz fills the 288-MB solid-state flash disk in 21 h, but this lifetime can be extended by reducing resolution or by recording only during times of interest. Low-power 3-V electronics allow a single half-AA-cell lithium battery to power the entire tag. The tags were deployed as seafloor recorders off Maui in March 2001, and initial deployments on free-ranging marine mammals and sea turtles are expected by fall 2001. [Work supported by ONR.]

11:00

4aAB9. Ship strike acoustics: It is all just shadows and mirrors. Joseph E. Blue (Leviathan Legacy, Inc., 3313 Northglen Dr., Orlando, FL 32806), Edmund R. Gerstein (Leviathan Legacy, Inc., 1318 S.W. 14th St., Boca Raton, FL 33486), and Steve E. Forsythe (Navy Undersea Warfare Ctr. Div. Newport, Newport, RI 02841)

Whales are vulnerable to collisions when near the surface and in shallow water. Here the physics of near-surface sound propagation may play a crucial role in their survival. Ships sufficient in size to mortally injury whales generate acoustic spectra dominated by very low frequencies. Since the ocean's pressure-release surface severely attenuates frequencies that are generated at distances less than a wavelength from the surface, whales may not detect low-frequency sounds generated by approaching ships. The Lloyd Mirror Effect predicts sound pressure levels at the surface approximate zero and empirical acoustic measurements support the prediction, while quantifying another important phenomenon associated with large ships, Acoustical Shadowing. The confluence of these acoustic propagation effects pose significant ecological consequences for marine mammals at the surface. While aerial and shipboard monitoring networks help navigate ships away from whales sighted on the surface, these programs are not reliable during conditions of poor visibility and darkness. Further understanding the acoustics of whale/ship collisions could augment protection efforts and result in more consistent and effective management strategies. To mitigate collision risks at times when surveillance programs are ineffective, ships could be projecting directional acoustic signals specifically designed to defeat the Lloyd Mirror Effect and Acoustical Shadowing.

THURSDAY MORNING, 6 DECEMBER 2001 ROOM 305, 8:00 TO 10:15 A.M.

Session 4aAOa

Acoustical Oceanography: Seafloor Acoustics

Dajun Tang, Chair

Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

Contributed Papers

8:00

4aAOa1. Fine-scale volume heterogeneity measurements in sand. Dajun Tang, Kevin Williams, Darrell Jackson, Eric Thorsos (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105), and Kevin Briggs (Naval Res. Lab., Stennis Space Center, MS 39529-5004)

During the High Frequency Sediment Acoustics Experiment (SAX99), fine-scale variability of sediment density was measured by an *in situ* technique and core analysis. The *in situ* measurement was accomplished by a new instrument that measures sediment conductivity. These measurements were conducted on a three-dimensional grid, hence providing a set of data suited for studies of sediment acoustic scattering. A three-dimensional matrix of sediment density measurements is obtained from the conductivity data through an empirical relationship. Based on the density data, a power spectrum of density variability is estimated. Cores were collected independently at the SAX99 site from which density variations were measured using gravimetric techniques. Volume heterogeneity power spectra estimated from cores are compared to the ones obtained from the *in situ* measurements. It is found that whereas the overall volume heterogeneity in the sediment is very low, and thus contributes little to backscatter, the surficial part of the sediment has a rather strong heterogeneity apparently due to bioturbation. To model such scatter correctly, the surface roughness and the volume heterogeneity need to be considered jointly. This depth-dependent heterogeneity is expected to be significant at higher frequency (greater than 100 kHz). [Work supported by ONR.]

8:15

4aAOa2. Coastal seabed tomography by inversion of drifting acoustic buoys data. Jean-Pierre Hermand (Dept. of Optics and Acoust., Université Libre de Bruxelles, av. F.-D. Roosevelt 50, CP 194/5, B-1050 Brussels, Belgium) and Frans G. J. Absil (Dept. of Sensor, Weapon, and Command Systems, Koninklijk Instituut voor de Marine, Postbus 10000, 1780 CA Den Helder, The Netherlands)

A buoy field, specifically designed for shallow-water acoustic tomography, was deployed in a complex coastal environment on the western Sicilian shelf (ENVERSE 97, Saclant Undersea Research Centre). Eight buoys drifted away from a fixed sound source receiving its repeated, broadband transmissions on their single hydrophone at fixed depths and known positions (DGPS). The 1-day run sampled, over kilometers, the acoustic impulse response (the Green's function) of the environment as a function of range and azimuth from the source. In this paper, the mapping of sediment properties and bottom types from the ''synthetic horizontal aperture'' measurements is investigated. Statistical analyses of the frequency-dependent, mode interference patterns associated with the buoy tracks allow to isolate range-azimuth sectors of limited acoustic variability and to define regions of similar bottom conditions. Average geoacoustic parameters of these regions are determined by maximizing the processing gain of a model-based matched filter receiver [J.-P. Hermand, IEEE J. Oceanic Eng. **24**, 41–66 (1999)]. The preliminary inversion results, including P-wave speed and thickness of the sediment cover, are congruent with the ground truth of a dense grid of seismic reflection profiles and sediment cores (ENVERSE 98). [Work supported by the Royal Netherlands Navy.]

8:30

4aAOa3. Sediment tomography in the East China Sea: Preliminary results. Gopu R. Potty, James H. Miller, Chuen-Song Chen (Univ. of Rhode Island, Narragansett, RI 02882), Colin J. Lazauski (Univ. of Rhode Island, Narragansett, RI 02882 and Naval Undersea Warfare Ctr., Newport, RI 02841), and Peter H. Dahl (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

This paper discusses the early results from the acoustic bottom interaction experiment conducted in May–June 2001 in the East China Sea as part of the ASIAEX-2001. Explosive sources (38 g and 1000 g) were deployed in the experimental area from a Chinese Research vessel (Shi Yan 2) during the experiment. These shots are deployed in circular pattern of radius 30 km and on radial tracks. The acoustic signals from these explosive sources were received at the vertical line array hung from R/V Melville near the center of the circular pattern. The sediment properties show a sharp contrast on either side of a sediment front. The presence of high-speed sand and low speed mud–sand on either side of the front has been confirmed by core data from ASIAEX-2000. Extensive coring to determine the sediment information in more detail is planned in August 2001. The time-frequency scalograms of the received explosive signals have been constructed using wavelet analysis. This analysis yields the dispersion characteristics of the acoustic signals and clearly shows the modal arrivals. The dispersion behavior of the acoustic modes from either side of this sediment front is presented. These modal dispersion diagrams form the basis for our sediment inversion technique. [Work supported by ONR.]

8:45

4aAOa4. Influence of water column statistics and environmental parameters on time-domain simulations of SWARM 95 experimental data. Scott Frank, William L. Siegmann (Dept. of Math. Sci., Rensselaer Polytechnic Inst., Troy, NY 12180), and Mohsen Badiey (Univ. of Delaware, Newark, DE 19716)

During the 1995 SWARM experiment acoustic data were collected at two different VLAs from an airgun source aboard the R/V CAPE HATTERAS. Intense nonlinear internal wave activity occurred during approximately 1 h of this data. The received signals contain significant variability, particularly of their intensity and time-frequency composition. Efforts to simulate the variability observed in the experimental data by using parabolic equation calculations and simplified analytic models have indicated the importance of bottom physical properties. Results from both methods show sensitivity to the structure of the bottom sound speed as well as the statistics of the internal waves that were present. Bottom properties are estimated from core data taken near the site and other previous studies in the area. To analyze the effects of the internal waves successfully, necessary modifications to these properties are described. Water column statistics are investigated using thermister data taken during the experiment. These statistical quantities will be compared with observed variability in the received signals. Analysis of these features provides relationships between internal wave variability and broadband interference patterns, intensity variations, and modal amplitude and arrival structure. [Work supported by ONR.]

9:00

4aAOa5. Suction cup effect coupling shear waves across unconsolidated smooth wet solids. Jacques R. Chamuel (Sonoquest Adv. Ultrason. Res., P.O. Box 81153, Wellesley Hills, MA 02481-0001)

Accurate modeling of the propagation of sound waves in unconsolidated marine sediments depends on understanding the interparticle contacts and coupling mechanisms of compressional and shear waves. The theory of elastic waves in unconsolidated marine sediments [M. J. Buckingham, J. Acoust. Soc. Am. **103**, 288–299 (1998)] depends on the grain roughness, where the shear speed goes to zero as the losses vanish. The current paper examines shear wave coupling phenomena between unconsolidated solids in the absence of roughness. Fundamental experimental results are presented, demonstrating the transmission of shear waves across smooth glass plates coupled with a localized capillary thin water film hermetically sealed between the plates with the suction cup effect. The glass plates were pressed against each other while immersed in distilled water until they became slip resistant. The diffraction pattern of Newton fringes was observed surrounding the water film. Results from experiments using seawater, microbubbles, and curved surfaces are discussed leading to acoustic characterization of saturated and partially saturated marine sediments. [Work supported by ONR.]

9:15

4aAOa6. Direct observation of the angle of intromission in marine sediments. Charles Holland (Appl. Res. Lab., Pennsylvania State Univ., P.O. Box 30, State College, PA 16804)

High porosity marine sediments like silty clays have the curious property that the speed of sound through its bulk medium is lower than that of the interstitial pore fluid. When a high porosity sediment is at the water sediment interface, classical theory predicts that there is an angle at which the reflection coefficient is zero, and there is total transmission of sound into the seafloor. This angle is called the angle of intromission and has not been directly observed at the seafloor to the author's knowledge. Data from a new measurement technique show this phenomenon with remarkable clarity. Measurements of the angle of intromission in the coastal region of Italy indicate that the properties of the surficial high porosity sediments are surprisingly constant over large areas. A simple, but robust inversion method is shown for which the sediment sound speed and density can be directly obtained.

9:30

4aAOa7. Normal-incidence measurements of the reflection coefficient of two sand sediments in the Gulf of Mexico. Eric Giddens, Thomas K. Berger, Iain Clark, and Michael J. Buckingham (Scripps Inst. of Oceanogr., Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0238)

In June 2001, the normal-incidence bottom-reflection coefficient, R, of two (20 m depth), sandy seabeds off Florida in the Gulf of Mexico (the SAX'99 site off Destin and another site off Panama City) was measured using tone bursts of 5, 10, 20, and 30 kHz. An omnidirectional source was suspended 2 m above the bottom and 1 m below a hydrophone, which detected the direct and reflected signals. R was determined with the sediment in its natural state and also after surface features, typically ripples of 0.5 m wavelength and 5 cm amplitude, had been smoothed. For any given run, values of R for the undisturbed sediment usually show higher variability than those for the smoothed bottom. The mean value of R for the natural sediment is -9.85 dB and for the smoothed bottom -8.0 dB, the latter being consistent with Rayleigh. Since the source–hydrophone pair was free to drift slightly relative to the seabed, the high variability of the natural-bottom returns may be due either to mild focusing and defocusing effects from the peaks and troughs of the ripple field or to scattering from small inhomogeneities such as shell fragments in the sediment. [Work supported by ONR.]

9:45

4aAOa8. Acoustic wave reflection from the transition layer of surficial marine sediment. Masao Kimura (Dept. of Geo-Environ. Technol., Tokai Univ., 3-20-1 Orido, Shimizu, Shizuoka 424-8610, Japan, mkimura@scc.u-tokai.ac.jp)

It has been recently reported that the porosity and the density, etc. in the surficial marine sediment vary largely with the depth. The characterisitics of acoustic wave reflection from such transition layers of the surficial marine sediment seem to be very different from that of uniform sediments. In this study, the incident angle dependence of the reflection coefficient for the transition layer of the surficial marine sediment model is calculated using OASES (Biot–Stoll model). The effects of the transition layer and the frequency dependence of the reflection coefficient characteristics are investigated. Next, the characteristics of the incident angle for the reflection coefficient are measured in water tank and *in situ*. Beach sand (fine sand) is used for water tank measurements, and *in situ* measurements are done in a very fine sand bottom. The operating frequency is 150 kHz. The incident angle are 0, 20, 50, and 60 deg. The measured results are compared with the calculated results using OASES (Biot–Stoll model), and the effects of the transition layer on the characteristics of the reflection coefficient are investigated.

10:00

4aAOa9. Observations of sediment shear velocity in the San Diego Trough. LeRoy M. Dorman and Allan W. Sauter (Scripps Inst. of Oceanogr., Univ. of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0220)

We report observations of shallow seafloor shear velocity structure in the San Diego Trough, whose depth is about 1100 m. Dispersed Scholte waves were generated using a new implosive source, described elsewhere in this meeting, and observed using ocean-bottom seismographs. Preliminary estimates of the surficial shear velocities are in the 16–25 m/s range, which is very low. The Scholte waves are observed in the 1–8 Hz frequency range and dispersion is well-developed after 100 m of propagation. Higher modes appear in addition to the fundamental mode, which exhibits a smooth increase in frequency from 2 to an Airy phase at about 3 Hz. The group velocity of the Airy phase approximates the surficial shear velocity. In the work area, the topographic relief is only a few meters. Geologically, the sediments are turbidites distributed by a channel and levee system. High-resolution multibeam bathymetry is available from NOAA and is being used to aid geological interpretation.

THURSDAY MORNING, 6 DECEMBER 2001 ROOM 305, 10:30 A.M. TO 12:00 NOON

Session 4aAOb

Acoustical Oceanography: Acoustic Remote Sensing

Jeffrey A. Nystuen, Chair
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

Contributed Papers

10:30

4aAOb1. The underwater sound field from Crater Lake, Oregon. Jeffrey A Nystuen (Appl. Phys. Lab., The Univ. of Washington, Seattle, WA 98105) and Robert Collier (Oregon State Univ., Corvallis, OR)

Breaking waves and precipitation on the ocean surface are primary sources of underwater sound in the frequency range from 1–50 kHz. By analyzing the underwater sound, precipitation can be detected and quantified, and wind speed can be measured. Acoustic Rain Gauges (ARGs) have been designed and deployed on oceanic moorings. An opportunity to deploy an ARG on a mooring in Crater Lake, Oregon allows a contrast between fresh water and salt water deployments. Furthermore, the possibility of acoustic snowfall detection is investigated. Acoustic signatures for drizzle and heavy rain are similar to those obtained in oceanic conditions. However, acoustic snowfall detection was not readily apparent. Most of the snowfall apparently occurred during high wind speed conditions of over 10 m/s. The marine acoustic wind speed algorithm is biased toward higher values relative to the surface anemometer when the wind speed is over 10 m/s. The spectral distortion of higher-frequency sound levels due to entrained ambient bubbles is not present in these fresh water data. This is consistent with the observation that whitecaps in fresh water are made of larger bubbles than in salt water and that the whitecaps do not persist as long in fresh water.

10:45

4aAOb2. Fermat's principle and acoustic tomography in a time-dependent ocean. Oleg A. Godin (CIRES, Univ. of Colorado at Boulder and NOAA/Environ. Technol. Lab., 325 Broadway, Boulder, CO 80305-3328) and Alexander G. Voronovich (NOAA/Environ. Technol. Lab., 325 Broadway, Boulder, CO 80305-3328)

Fermat's principle states that, for nondispersive waves propagating between two given points, travel time is stationary with respect to small variations of the wave propagation path around the actual path (i.e., ray) taken by the wave. This fundamental principle governing short wave kinematics was first established by seventeenth century French mathematician Pierre de Fermat. The principle is of particular importance for acoustic tomography because it allows one to neglect perturbations in ray geometry when performing a linear inversion of acoustic travel times for sound speed variations. Derivations of Fermat's principle available in the literature does not apply to acoustic waves in media with time-dependent sound speed. Moreover, the very formulation of the principle in this case is not obvious as travel time along a trial ray ceases to be a single-valued functional of the trial ray geometry. In this paper, the Fermat's principle is extended to media with time-dependent parameters. We consider acoustic waves in the inhomogeneous, moving or quiescent fluid. Formulations of the principle in terms of travel time and eikonal are compared. Applications of the established principle to ocean acoustic tomography and modeling wave propagation in weakly nonstationary, inhomogeneous ocean are discussed. [Work supported by ONR.]

11:00

4aAOb3. Observations of intensity fluctuations during the summer New England shelf break Primer study. Andrew J. Fredricks, John Colosi, and James F. Lynch (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

We present intensity fluctuation statistics of point-to-point transmission data taken during the summer shelfbreak Primer experiment. We focus on broadband 400 Hz transmissions from two bottom mounted sources on the continental slope to a vertical line array on the shelf, at distances of 60 and 42 km. The data were collected over 11 days in late July and early August of 1996. Single phone, time averaged, and time and depth average intensity statistics are examined. Both tracks showed a strong M2 tidal effect on intensity, as well as mesoscale induced variations. [Work supported by ONR.]

11:15

4aAOb4. Observing barotropic currents and tides in the Kuroshio Extension region from long-range reciprocal acoustic transmissions. Gang Yuan, Humio Mitsudera (IPRC-SOEST, Univ. of Hawaii, 2525 Correa Rd., Honolulu, HI 96822), H. Fujimori, T. Nakamura, Y. Yoshikawa, and I. Nakano (Japan Marine Sci. and Tech. Ctr., Yokosuka 237, Japan)

Travel times of reciprocal acoustic transmissions from a 5-transceiver acoustic tomography array over a 1000-km-scale domain in the Kuroshio Extension (KE) region during summer 1997 are used to investigate the barotropic tidal currents and range-averaged large-scale low-frequency currents as well as their relative vorticity. Inverse techniques are employed to estimate the low-frequency range-and-depth averaged (barotropic) currents every 3 or 6 h from the differential acoustic travel times. Eight major tidal constituent amplitudes and phases derived from the high-frequency acoustic travel times agree well with those found from the TOPEX/POSEIDON tidal model TPXO.5 [Egbert *et al.*, 1994]. The results show that the range-averaged barotropic currents along the section in the KE recirculation gyre flowed westward in about 4 cm/s with an uncertainty of 0.8 mm/s; this is consistent with the result of the ADCP survey on WOCE P-14N experiment, and comparable with the surface geostrophic current velocity 5 m/s determined from the SSHA of the TOPEX/POSEIDON altimeter, indicating that barotropic currents are dominant in the Kuroshio Extension recirculation gyre. Estimated daily averaged relative vorticity over a 155 000 km^2 area in the KE recirculation region has a near-zero mean (1.84×10^{-7} s^{-1}).

11:30

4aAOb5. A multifrequency scintillation probing of current in a stratified ocean. Iosif Fuks, Mikhail Charnotskii, and Konstantin Naugolnykh (NOAA/Environ. Technol. Lab/Zeltech/Univ. of Colorado, 325 Broadway, Boulder, CO 80305)

Transverse flow of inhomogeneous fluid produces fluctuation of the acoustic signal passing through it. These fluctuations vary with the CW signal frequency change due to variation of the Fresnel zone and diffraction angles. Respectively, the fluctuations of signals at two different frequencies are coherent in a low frequency range of the spectrum and noncoherent in the high-frequency band. The frequency cutoff of the correlation function of the signal fluctuation depends upon the flow velocity at given fine structure of the flow. Measurement of the cutoff frequency allows one therefore to determine the flow velocity of the fluid. The principles of this approach in application to the probing of current in the homogeneous in average ocean were considered previously [J. Fuks, K. Naugolnykh, and M. Charnotskii, J. Acoust. Soc. Am. **109**, 2730–2738 (2001)]. In a stratified ocean the sound field has a modal structure connected with the fact that rays tend to bend toward region of smaller sound speed. The features of the multifrequency method of current measurements in a stratified ocean when the fine structure of the flow is determined by the realistic spectrum of internal waves and the sound field is presented as a modal set are considered in the present paper.

11:45

4aAOb6. Energy: Converting from acoustic to biological resource units. Kelly J. Benoit-Bird and Whitlow W. L. Au (Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734)

Acoustic backscattering strength is often used as an index of biomass; however, the relationship has not been directly validated. Relationships were investigated between acoustic cross section at 200 kHz and measured values of length, biovolume, dry weight, ash-free dry weight, and caloric content of the same individual specimens. Animals were part of the Hawaiian mesopelagic boundary community and included shrimps, squids, and myctophid fishes. The strong linear relationships found between all the variables measured make it possible to approximate any one variable from the measured values of others within a class of animals. These data show that for these midwater animals, acoustic scattering can be used as an index of biomass. Dorsal-aspect acoustic cross section at 200 kHz predicted dry weight and ash-free dry weight at least as well as did body length, a standard predictor. Dorsal-aspect acoustic cross section at 200 kHz was also a strong predictor of total caloric content. The relationships between dorsal-aspect acoustic cross section and both weight and caloric content of Hawaiian mesopelagic animals were linear and additive. Consequently, it is possible to directly convert acoustic energy from these animals to organic resource units without having knowledge of the size distribution of the populations being studied.

Session 4aEA

Engineering Acoustics: Ferroelectric Single Crystals: From Manufacture to Devices

Elizabeth A. McLaughlin, Chair
Naval Undersea Warfare Center Division Newport, Code 2132, 1176 Howell Street, Newport, Rhode Island 02841

Chair's Introduction—8:10

Invited Papers

8:15

4aEA1. Single crystal piezoelectrics for advanced acoustic transducers. Wesley Hackenberger, Paul W. Rehrig, Ming-Jen Pan, and Thomas Shrout (TRS Ceramics, Inc., 2820 E. College Ave., State College, PA 16801, wsh@trsceramics.com)

Single crystal piezoelectric materials based on $Pb(Mg_{1/3}Nb_{2/3})_xTi_{1-x}O_3$ (PMN-PT) exhibit extraordinarily high piezoelectric properties ($d_{33}>2000$ pC/N; $k_{33}>90\%$), which are being utilized to produce transducers with revolutionary performance enhancements in sensitivity and bandwidth. Crystals of PMN-PT are currently being produced in sizes of >2.54 cm in diameter by >7 cm in length. Examples of transducer enhancements demonstrated to date include medical ultrasound transducers with bandwidths in excess of 100%, similarly broadband Tonpilz transducer elements for forward look sonar (with the additional expectation of significantly higher source level over ceramic based transducers), and vibration and vector sensors with sensitivity levels approaching the theoretical noise floor. A number of issues need to be worked out to successfully commercialize crystals for these applications, the most important of which are reducing property variations that result from compositional gradients and increasing crystal size to reduce cost. Recent advances in crystal growth techniques are expected to help solve these problems in the coming year. Single crystal piezoelectric technology is continuing to develop with new materials that exhibit broader operating temperature ranges and larger electric field induced strains than PMN-PT. Thus single crystal piezoelectrics are expected to be as broadly applicable as ceramics.

8:45

4aEA2. Characterization of piezoelectric single crystals for sonar and actuator applications. Harold C. Robinson, Elizabeth A. McLaughlin, Lynn Ewart, and James M. Powers (NUWC, 1176 Howell St., Newport, RI 02841, robinsonhc@npt.nuwc.navy.mil)

Piezoelectric single crystals show great promise for sonar and actuator applications due to their very high piezoelectric constant and electromechanical coupling factor. However, in order to accurately assess their potential, the effect of operating conditions (i.e., applied electric field, mechanical loading, and temperature) on their material properties must be determined. Furthermore, the mechanical strength of these materials in compression and tension must be known to avoid mechanical failure in transducers. A brief overview of the various methods used to characterize these crystals will be presented. In particular, material properties obtained using the NUWC Newports Stress Dependent Electromechanical Characterization System (SDECS) will be detailed. The variability of these properties under applied mechanical prestress, high electric fields, and varying temperatures shall be examined. An enhancement of the piezoelectric response with the application of prestress in certain crystals will be shown. In addition, results from compression strength testing conducted at NUWC will be presented. It shall be demonstrated that these crystals are much stronger than was assumed previously. Finally, future challenges in the measurement and analysis of these materials will be identified, including the identification of material flaws and quantification of the processes that lead to cracking in piezocrystals. [Work sponsored by DARPA.]

9:15

4aEA3. Characterization of the nonlinear properties of PZN–PT. Christopher S. Lynch (The GWW School of Mech. Eng., The Georgia Inst. of Technol., Atlanta, GA 30332-0405)

This presentation will focus on the electro-mechanical properties of single crystals of relaxor ferroelectric PZN–PT. The characterization work begins with measurement of the drive capability of the material as a function of preload. These measurements were made by loading specimens at constant stress levels and then cycling the electric field while monitoring the resulting strain. Results to be presented are curves for stress/strain at various electric field levels, and strain/electric field and electric displacement/electric

field at various constant stress loads. A second set of experiments was performed by placing the specimens in bending. In this arrangement, the tensile and compressive stress/strain behavior are determined simultaneously. This clearly shows the effects of material nonlinearity. The final set of tests takes a close look at fatigue of the material. Small cracks are intentionally placed in specimens and then the electric field is cycled. Observation of fatigue behavior will be discussed.

9:45–10:00 Break

10:00

4aEA4. Development and commercialization of high-performance sensorics utilizing single piezoelectric crystals. Paul A. Wlodkowski (Wilcoxon Res., Inc., 21 Firstfield Rd., Gaithersburg, MD 20878)

Wilcoxon Research, Inc., has two focused research efforts on developing high-performance sensorics with single piezoelectric crystals. The first program belongs to the emerging class of directional acoustic transducers, also known as vector sensors. The second program involves the development of mesoscale accelerometers. Recognizing both the exceptional electromechanical properties and the limitations of the new piezoelectric material, the company has identified the aforementioned areas as those most promising for commercialization. Extensive materials characterization, prototype fabrication and testing have supported this direction. Wilcoxon Research has successfully demonstrated the unprecedented performance of both classes of transducers. This has been realized through a synergy of the electromechanical properties of single piezoelectric crystals, a novel method of strain amplification in the material, and low-noise signal conditioning customized for the PMNT and PZNT crystals. Typically, the signal-to-noise ratio of the single piezoelectric-based sensors is 15 dB higher via-à-vis conventional transducers utilizing polycrystalline, PZT ceramic materials as the sensing element.

10:30

4aEA5. Application of single crystal relaxor ferroelectric materials to medical imaging arrays. M. Zipparo and C. Oakley (Tetrad Corp., 357 Inverness Dr. S., Ste. A., Englewood, CO 80112)

Single crystal relaxor ferroelectrics have been shown to possess extremely high electromechanical coupling coefficients, in some materials approaching 95% in length extensional mode. This high coupling allows the design of electroacoustic transducers with better sensitivity or bandwidth than can be achieved with conventional piezoceramic materials. These improved properties come at the expense of lower depoling temperatures, increased material fragility, poorer property uniformity, and increased cost. These issues preclude the adoption of single crystals as a direct drop-in replacement for conventional piezoceramics. Most issues can be addressed through refined crystal growth and transducer design or processing methods. This paper reviews the application of single crystals to transducers for medical imaging, including: characterization of material properties including the complete matrix of elastic, dielectric, and piezoelectric coefficients; design of high bandwidth transducers; modeling of transducer operation, including 1-D and finite element approaches; development of alternate fabrication processes; and measurement of the transducer performance improvement which has been obtained to date. Progress toward the goal of fabricating a single transducer which can operate effectively over the range of multiple conventional probes will be described.

11:00

4aEA6. Recent progress in single crystal naval sonar transducer development. James M. Powers (NUWC, 1176 Howell St., Newport, RI 02841, powersjm@npt.nuwc.navy.mil)

Single crystals of lead magnesium niobate/lead titanate (PMN-PT) and lead zinc niobate/lead titanate (PZN-PT) are currently under investigation as high-performance piezoelectric transducer materials. In particular the high-strain ($>1.0\%$) and high-coupling factor ($k_{33}>0.90$) potentially allow the design of sonar transducers with greatly increased power density and bandwidth. In this talk the properties of single crystal material will be discussed as they pertain to sonar transducers, and some ongoing single crystal transducer developments will be reviewed. [Work supported by DARPA and ONR.]

Session 4aMU

Musical Acoustics and Education in Acoustics: Musical Acoustics in Teaching Science, Engineering and Technology

Robert D. Collier, Cochair
Thayer School of Engineering, Dartmouth College, Cummings Hall, Hanover, New Hampshire 03755-8000

E. Carr Everbach, Cochair
Department of Engineering, Swarthmore College, Swarthmore, Pennsylvania 19081-1397

Chair's Introduction—8:00

Invited Papers

8:05

4aMU1. Musical acoustics in the science classroom. Uwe J. Hansen (Dept. of Phys., Indiana State Univ., Terre Haute, IN 47809)

The value of hands-on laboratory centered teaching of science has led to continued workshops for elementary and secondary school teachers. Emphasis for teachers in the primary grades has been on the use of music as a tool to introduce science concepts, and for teachers at the secondary level to use computer based instrumentation to perform relatively sophisticated measurements. In a recent issue of ''The Hoosier Science Teacher'' 13 activities are discussed for use in a program which uses music to introduce science concepts [U. J. Hansen, The Hoosier Science Teacher, Volume XXVI, Number 1, 4–14 (2000)]. Increasing accessibility to sophisticated computer software enhances the ability to perform rather complex experiments at the high school level. Recent workshops at regional meetings of the AAPT have introduced these concepts to High School Physics teachers. Both the role of music in the elementary science curriculum and the enhancement of the High School Physics laboratory will be discussed with specific examples.

8:35

4aMU2. Musical acoustics in teaching engineering to first-year undergraduates. E. Carr Everbach (Eng. Dept., Swarthmore College, Swarthmore, PA 19081)

Since most high school programs lack any introduction to the fields of engineering, first-year undergraduates often take an ''inspirational overview'' course to see if they wish to pursue the field. Musical acoustics provides accessible examples of phenomena and concepts that can draw such students into further engineering applications. Furthermore, the programming environment Matlab, a long time staple of engineering analysis, has evolved such sophisticated capabilities that nearly real-time analysis and synthesis of musical acoustics demonstrations is possible in the lecture hall. Examples of linear and nonlinear systems in musical acoustics will be demonstrated and the Matlab source code (m-files) freely available via anonymous ftp at ftp://fubini.swarthmore.edu

9:05

4aMU3. Descriptive acoustics of music and speech. Timothy W. Leishman (Dept. of Phys. and Astron., Brigham Young Univ., N335 ESC, Provo, UT 84602)

Approximately 250 nonscience undergraduate students enroll each year in a Brigham Young University physics course entitled Descriptive Acoustics of Music and Speech. This general education service course, which has been taught for more than 30 years, provides an introduction to physical concepts underlying the production and perception of music and speech. Participating students receive theoretical grounding and hands-on experience in principles of general acoustics, human hearing, architectural acoustics, sound reproduction, human voice, and musical instrument acoustics. The course was originally formulated with strong encouragement and support from the faculty of other departments who wanted to have a course that encountered their disciplines from the physical science viewpoint. Sound, as a familiar area of experience to students, allows them to view aspects of their disciplines in contrasting and complementary ways. The course has become a requirement for many nonscience students and a preferred elective for others. This paper presents an overview of techniques that have been successfully used over the years to teach acoustics to these students. It also discusses new approaches that are currently under development to enhance their learning experience.

9:35

4aMU4. Musical instruments as tools for teaching acoustics to engineering students. Daniel A. Russell (Sci. and Mathematics Dept., Kettering Univ., Flint, MI 48504, drussell@kettering.edu)

Kettering University offers an undergraduate minor in acoustics anchored by two lecture courses and a laboratory project course taught by physics faculty. While many of the students who take these courses are more interested in the engineering and noise control applications of acoustics, most are familiar with musical instruments and quite a few play an instrument. We have found that in several instances a discussion of the acoustics of a musical instrument can be very effective at facilitating an understanding of the fundamental

concepts behind the engineering applications of acoustics our students encounter at their co-op work experiences. Examples include: the physics of guitar pickups, hose trumpets, ocarinas, low-frequency models of acoustic guitars, sound radiation from drum membrane modes, radiation from a tuning fork, musical scales and intervals, and audio demonstrations of pitch and timbre. This talk will discuss and demonstrate how several such examples are used as classroom demonstrations, homework projects, and laboratory experiments.

10:05–10:20 Break

10:20

4aMU5. Learning about waves with musical acoustics. Thomas D. Rossing (Phys. Dept., Northern Illinois Univ., DeKalb, IL 60115)

Sound waves, water waves, quantum waves, light waves, and other types of electromagnetic waves have much in common that can be understood by doing experiments in wave propagation with one or more of these types of waves. Sound waves are particularly useful because of their convenient wavelengths, and sound waves associated with music or musical instruments are attractive to students. We describe experiments in musical acoustics that have applications to other areas of physics and to engineering.

10:50

4aMU6. The role of musical acoustics in teaching engineering. Robert D. Collier (Thayer School of Eng., Dartmouth College, Hanover, NH 03755)

A new course, Technology of Sound, is being offered at Dartmouth College to introduce students in the liberal arts to science and engineering. The course is designed to demonstrate how the science of sound is being applied to enhance the quality of life through music making and listening. The fundamentals of acoustics are studied to understand the relationships between the vibrations and sounds of string, wind, reed, brass, and percussion instruments. Associated properties and design criteria of materials, structures, and control mechanisms including musicians are studied. Key features of the course include lecture demonstrations of individual instruments, term projects and laboratory investigations, and studies of architectural acoustics. Specific examples of student projects and the development of a web site as a teaching tool and resource will be discussed as well as student comments and evaluations. A visit to Symphony Hall in Boston is a highlight at the end of the course.

Contributed Paper

11:20

4aMU7. Radiation patterns of novel one- and three-tine tuning forks. Earl D. Blodgett, Lowell I. McCann (Dept. of Phys., Univ. of Wisconsin, River Falls, River Falls, WI 54022), and Chelsea Anderson (Hudson High School, Hudson, WI 54016)

Inspired by recent results [D. A. Russell, Am. J. Phys. **68**, 1139–1145 (2000)] showing that the traditional two-tine tuning fork behaves as a linear quadrupole source, novel one-tine and three-tined tuning forks were constructed. The three identical tines on the three-tined fork were equally spaced around a circle and oriented so that each tine oscillated radially when exciting the fundamental mode. The angular dependence of the radiated sound from these forks was investigated as a function of distance from the symmetry axis of each fork. In the near field, the three-tined fork has three major maxima, one adjacent to each tine, and three minor maxima, one opposite each tine. The minor maxima disappear in the far-field measurements. These results were compared to a simple model wherein each tine behaves as an independent point dipole source [P. M. Morse and K. U. Ingard, *Theoretical Acoustics*, 2nd ed. (Princeton U.P., Princeton, NJ, 1986), p. 313]. The simple model shows good, but not perfect, agreement with the measured radiation patterns.

11:35–12:00
Panel Discussion

Session 4aNS

Noise: Community Noise: Aircraft and Other Transportation Sources

Nancy S. Timmerman, Cochair
Noise Abatement Office, Massport, One Harborside Drive, East Boston, Massachusetts 02128-2909

Bennett M. Brooks, Cochair
Brooks Acoustics Corporation, 27 Hartford Turnpike, Vernon, Connecticut 06066

Chair's Introduction—8:00

Invited Papers

8:05

4aNS1. Lateral attenuation of aircraft sound levels due to engine installation effects: Wallops study. Gregg G. Fleming (Volpe Natl. Transportation Systems Ctr., Environ. Measurement and Modeling Div., Kendall Square, Cambridge, MA 02142), Kevin P. Shepherd (NASA Langley Res. Ctr., Hampton, VA 23681-0001), and David A. Senzig (Senzig Eng., 7 Everett St., Boston, MA 02130)

In September of 2000, the National Aeronautics and Space Administration's (NASA) Langley Research Center, with assistance from the John A. Volpe National Transportation Systems Center conducted a controlled aircraft noise measurement study at NASA's Wallops Island Flight Facility on the northeast coast of Virginia. The broad objective of the study was to obtain the data necessary to support an update/replacement to SAE AIR 1751, which addresses the lateral attenuation of aircraft noise. The specific focus of the study was the engine/installation effects, or directivity characteristics of various types of civil transport aircraft. Included in the study was a 767-400, a DC-9, a Falcon 2000, and a King Air turboprop. Measurements were conducted with a 20 microphone array deployed on a combination of two construction cranes and a set of 30 ft. poles. Precision aircraft tracking data were collected using a differential global positioning system, as well as a video tracking system. This presentation will overview the results of the study.

8:30

4aNS2. Measured versus modeled aircraft noise levels at Hull, MA. Paul Schomer (Shomer and Assoc., 2117 Robert Dr., Champaign, IL 61821, schomer@staff.uiuc.edu)

Boston's Logan Airport maintains an unattended noise monitor on the Hull peninsula by the Hull High School. They also produce yearly DNL contours for aircraft operations at the airport using INM. These contours include Hull. This paper compares the measured levels at Hull with the DNL predictions. This comparison is done in two steps. First, the true measured airport noise is separated from other monitor-measured noise—notably wind and wave noise. Second, these cleaned data are compared with the DNL predictions. The results show that INM underpredicts at Hull by at least 3 decibels. Of these 3 decibels, at least one can be attributed to the over water (hard surface) situation and the remaining difference is thought to result from systematic sideline biases in INM that relate to other propagation and engine shielding effects.

8:55

4aNS3. Prediction of aircraft flight tracks from noise measurements. Bruce Ikelheimer, Kenneth Plotkin (Wyle Labs., 2001 Jefferson Davis Hwy., Ste. 701, Arlington, VA 22202, bikelheimer@arl.wylelabs.com), and Timothy Lavallee (Geo-Marine, Inc., 11846 Rock Landing Dr., Ste. C, Newport News, VA 23606)

A method has been developed to estimate an aircraft's flight path based on noise measurements at a number of positions within a study area. Two-second time histories of A-weighted sound levels were recorded on 38 digital sound level meters (SLMs) distributed by Geo-Marine Inc. through a mountainous region of approximately 400 square miles. Six of these SLMs recorded one-third octave spectra. Aircraft pass through this area at low altitude along straight tracks (no turns). Determining the aircrafts position from SLM data is difficult because of the complexity of sound propagation over precipitous terrain and the size of the data matrices. The computational core of Wyle Laboratories' noise simulation model NMSIM is used to perform the propagation calculations. A least mean square method is used to curve-fit the flight track to the SLM data. The accuracy of this approach depends on the time intervals of the SLM data, the speed of the aircraft, and the accuracy of the NMSIM propagation algorithms. The NMSIM predicted flight path will be compared to available radar flight track data. Once a flight track has been established, NMSIM modeled time history and spectral data will be compared with measured data.

9:20

4aNS4. Highway traffic noise measurements compared to predictions from FHWA's Traffic Noise Model. Judith L. Rochat and Gregg G. Fleming (U.S. DOT, Volpe Natl. Transportation Systems Ctr., Cambridge, MA 02142)

In 1998, the United States Federal Highway Administration (FHWA) released a new tool for highway traffic noise prediction and noise barrier design, the Traffic Noise Model (TNM). In order to assess the accuracy and make recommendations on the use of TNM for the FHWA, the Volpe Center Acoustics Facility performed numerous roadside measurements, obtaining over 100 h of traffic noise data from highways around the country. For each site, acoustical, meteorological, and traffic data were collected simultaneously throughout the measurement period. Spectrum analyzers were used to collect 1/3-octave band *A*-weighted equivalent sound levels, and the microphones were deployed at distances from 50 to 1300 ft from the roadway and at two heights off the ground, the number of microphones used being site dependent. In comparing the measured and TNM-predicted sound levels, results indicate that TNM is, on average, within 1 dB of the measured data for all types of sites combined. Grouping the data by site types shows how well TNM is performing for open (no barrier), acoustically soft ground sites; open, acoustically hard ground sites; and sites with noise barriers.

9:45

4aNS5. How can we preserve and promote compatible land use in the age of shrinking aircraft noise contours? Deborah W. Murphy (URS Corp., 7650 W. Courtney Campbell Cswy., Tampa, FL 33607, deborah_murphy@urscorp.com)

Technological advances in source noise reduction combined with legislated or economics-driven changes in fleet mix composition have created the age of shrinking noise contours at many airports, even though the number of operations continues to increase. All too often, the majority of noise complaints come from residents located outside the airports Day–Night Average Sound Level (DNL) 65-dB noise contour, which is the threshold of compatibility established in the FAAs land use compatibility guidelines. If the local community feels strongly that their threshold of compatibility is less than DNL 65 dB, what can they do about it? The FAA will recognize a locally established standard and approve mitigation measures based on that standard. However, with that comes a responsibility on the part of the local jurisdictions to implement measures to prevent development of new noncompatible land uses in areas exposed to noise levels above their threshold of compatibility. This is crucial since the FAA will not approve or fund remedial mitigation of new noncompatible land uses built after October 1, 1998. This paper will explore ways the airport operator, the surrounding community, and the local jurisdictions may all be able to get what they want, with the blessing and funding of the FAA.

10:10–10:20 Break

10:20

4aNS6. Comparison of measured versus modeled aircraft noise levels. Mary Ellen Eagan (Harris Miller Miller & Hanson, Inc., 15 New England Executive Park, Burlington, MA 01803, meagan@hmmh.com)

The Integrated Noise Model (INM) is the primary tool used for predicting aircraft noise in the vicinity of civil airports in the U.S. and throughout the world. The INM contains a substantial database of more than 200 aircraft types for use in preparing noise contours. Validation of aircraft noise data in the database under field conditions has shown consistent underprediction for modeled versus measured data. Case studies of such observations will be presented.

Contributed Papers

10:45

4aNS7. The influence of a quiet facade on road traffic annoyance. Truls Gjestland (SINTEF Telecom and Informatics, N-7465 Trondheim, Norway) and Bard Stofringsdal (Norwegian Univ. of Science and Technol., N-7491 Trondheim, Norway)

A situation (1) with equal traffic noise on all four sides of a residence, has been compared with a situation (2) with a high level on one side and low on the other sides. A laboratory experiment has been conducted to have the test persons adjust the high or the low noise level in situation (2), so that the situation is assessed ''equally annoying'' as situation (1). The results indicate that the annoyance caused by road traffic noise can be reduced if at least one quiet side of the house can be achieved.

11:00

4aNS8. Noise pollution in the city of Curitiba, Brazil. Fabiano B. Diniz and Paulo H. T. Zannin (Environ. Acoust. Lab., Federal Univ. of Parana)

A study carried out in 2000 in the city of Curitiba, Brazil, showed that the noise levels generated by the traffic have decreased when compared to the levels obtained in a similar survey conducted in 1992. Noise levels were measured in 350 points spread through the city considering the same criterium adopted in 1992. A study conducted by Maschkle showed that levels higher than 65 dB(A) are likely to cause health disturbances. The current survey showed that 80.6% out of the measured points presented noise levels over 65 dB(A), whereas in the previous survey 93.4% out of the measured points presented levels over that value. In spite of this, the 2000 levels are still too high considering the local legislation, putting the city in a high risk category to health impairments. Among the possible reasons that caused such a reduction on the noise levels some can be cited: better conditions of the new road vehicles circulating in the Brazilian cities and in Curitiba as well, installation of electronic speed controllers, and the establishment of speed limits for some urban streets and avenues.

11:15

4aNS9. Social survey carried out in the city of Curitiba, Brazil. Fabiano B. Diniz and Paulo H. T. Zannin (Environ. Acoust. Lab., Federal Univ. of Parana)

As time passes, urban noise is more and more present in big cities inhabitants. Some studies have already quantified noise levels in the city of Curitiba, but they did not have the scope of analyzing the annoyance caused by these noises in its inhabitants. People are continuously exposed to noises, at any time of the day. When they sleep, talk, or even when they rest, the noise is often present. These factors contribute to health impairments and sleep disturbances, and can cause nervousness, headaches and

behavioral disorders. In the sense of seeking for people's annoyance and a reaction against the urban noise, a questionnaire was elaborated, intended to be distributed among the inhabitants of the city and returned for further analysis. By analyzing the answers it was possible to gather information on how the inhabitants of Curitiba react toward the urban noise. According to the answers, the traffic is the main noise source, which bothers more than 70% out of the interviewed people. Also, 32% out of the respondents feel annoyed by the noise in his/her street.

THURSDAY MORNING, 6 DECEMBER 2001 ROOM 222, 9:00 TO 10:30 A.M.

Session 4aPA

Physical Acoustics: Bubble Acoustics

Kerry W. Commander, Chair
Coastal Systems Station, Code R21, Panama City, Florida 32407

Contributed Papers

9:00

4aPA1. Further investigations on the effects of salts on SBSL. Anthony Khong, Elizabeth Doschek, Robert Apfel (Dept. of Mech. Eng., Yale Univ., 9 Hillhouse Ave., New Haven, CT 06520, anthony.khong@yale.edu), and Jeffrey Ketterling (Riverside Res. Inst., New York, NY 10036)

This current report extends earlier work done introducing lithium halide salts in the SL medium. A positive relationship between increasing salt concentrations in solution and bubble dynamics was noted. The radius versus time plots showed near perfect inelastic collapse of the bubble in lithium chloride solution as the salt concentration was increased to 3%. It appears that damping in the bubble is appreciably reduced in a solution with a high-salt saturation than in pure water. This observation seems in line with theories that considered the presence of water vapor molecules in the bubble as attenuating its collapse energy. However, any correspondence between the experimental and the theoretical C curves in the phase diagram plots could not be established. In this report, the following is examined: the effects of further increase in salt concentrations, gas saturations, and gas ratios in solution; of counterions other than lithium; and of lowering the temperature have on bubble dynamics. The aim, ultimately, is to have conditions streamlined for studies with salts in deuterated water, from which it is hoped to observe interesting products brought about by SL in this case. [Work supported by DARPA through a subcontract from University of Washington.]

9:15

4aPA2. A further validation of the dissociation hypothesis. Anthony Khong, Ning Xu, and Robert Apfel (Dept. of Mech. Eng., Yale Univ., 9 Hillhouse Ave., New Haven, CT 06520, anthony.khong@yale.edu)

The Dissociation Hypothesis states that in an air bubble or, more generally, a gas bubble comprised of a noble gas and a secondary gas molecule, nitrogen for instance, all the nitrogen is eventually removed by rectified diffusion in which water-soluble nitrogen species are formed. Only the noble gas would remain in the bubble. Hence, a final state exists in which, for a given gas mixture at a given gas saturation, the bubble content should consist predominantly, if not entirely, of the noble gas. An experiment was designed to examine this question. Keeping the final argon content constant, measurements were made of the bubble intensities and radii for various gas mixture permutations of argon-nitrogen under identical conditions of driving frequency, pressure amplitude and data acquisition time. It was found that between the cases 1 : 99 Ar : N2 at 20% saturation and pure Ar at 0.2% saturation, not only were bubble intensities noticeably less in the latter, so were the corresponding Ro, ambient radii, values. Also, Ro and bubble intensity became gradually smaller over time. Reconciliation of these data in light of DH will be discussed. [Support from DARPA subcontracted through University of Washington.]

9:30

4aPA3. Laser induced cavitational luminescence in a bipolar acoustic pulse wave. Georgii N. Sankin (Lavrentyev Inst. of Hydrodynamics, 15, Lavrentyev Str., 630090 Novosibirsk, Russia), Robert Mettin, Olgert Lindau, Bernhard Wolfrum, Werner Lauterborn (Drittes Physikalisches Inst., Goettingen, Germany), and Vyacheslav S. Teslenko (Lavrentyev Inst. of Hydrodynamics, Novosibirsk, Russia)

Laser induced bubble dynamics (200–300 μs life time) under the influence of a convergent bipolar acoustic pulse (100 bar, 2 μs) in water is investigated experimentally. The amplitude of both an emitted shock and luminescence from the bubble collapse are recorded by a pvdf pressure probe and a photomultiplier tube (which integrates the spectral range from 260 to 530 nm). Different delay time (phase) between collapse and pressure pulse is found to influence the luminosity. For certain phases the average amplitude of the luminescence signal is increased by an order of magnitude in comparison with the average signal from an undisturbed collapse (i.e., outside the additional pressure pulse). Photorecordings show that bubbles re-expand after the first collapse nonspherically if an acoustic pulse was applied. They form a nose on the shadow side of the pulse wave which is possibly indication for jetting. Results for a collapse of a laser cavitation bubble and a bubble produced by shock cavitation are compared. [Work supported by DAAD and RFBR.]

9:45

4aPA4. Measurements of the frequency- and size-dependent response of single, acoustically driven bubbles suspended in a polymer gel. Ryan D. McCormick and Ronald A. Roy (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215)

Sound propagation through bubbly liquids is a topic of interest to a variety of communities. However, characterization of the bubbly medium itself has proven difficult as bubbles either move with a flow or rise too quickly for precise observations. To bypass this problem, a gel has been devised using the food-thickening polymer xanthan gum to hold bubbles stationary pending dissolution [Hwang *et al.*, J. Atmos. Ocean Tech. **12**, 1289–1302 (1995)]. The gel has the same salinity as seawater, but possesses a much higher viscosity and is probably more elastic. Using this material, it is possible to unobtrusively ''levitate'' bubbles without reliance on an acoustic standing wave. The frequency-dependent driven response of a single bubble of known size can then be measured via laser scattering. Experimental results will be presented for an acoustically forced bubble driven at frequencies spanning the fundamental resonance frequency, and compared with model predictions. [Work supported by ONR.]

10:00

4aPA5. Giant acoustic stop bands for cubic arrays of air bubbles in water. M. S. Kushwaha (Inst. of Phys., Univ. of Puebla, P.O. Box J-45, Puebla 72570, Mexico), B. Djafari-Rouhani, and L. Dobrzynski (Dept. of Phys., Univ. of Science & Technol., Lille-I, 59655 Villeneuve D'Ascq, Cedex, France)

A systematic and extensive evidence is presented for the existence of complete, multiple, huge stop bands in the band structure for cubic arrays of air bubbles in water. This has been demonstrated by investigating face-centered cubic (fcc), body-centered cubic (bcc), and simple-cubic (sc) arrangements using the Fourier series expansion methodology, which does not require matching of the messy boundary conditions usually needed for the inhomogeneous composite systems. The lowest stop bands are largest for a volume fraction $f \leqslant 10\%$, with a gap/midgap ratio of ≈ 1.8 for all the three geometries. Surprisingly interesting but rigorously justifiable is the fact that the low-frequency flat passbands for the perfectly periodic systems correspond to the discrete modes of a *single* bubble. This is an artifact of the low filling fraction and huge density contrast in air and water. It is stressed that such a simple inhomogeneous system as made up of air bubbles in water exhibits the widest stop bands ever reported for phononic as well as photonic crystals. Some conclusive remarks will also be made on the reverse situation, where one can investigate the periodic system of water balloons in air.

10:15

4aPA6. A model for the ultrasonic modulation of optical path in acousto–photonic imaging. Sebastien Manneville, Ronald A. Roy (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215), Charles C. DiMarzio (Dept. of Elec. and Computer Eng., Northeastern Univ., Boston, MA 02215), and David A. Boas (NMR Ctr., Massachusetts General Hospital, Charlestown, MA)

Diffusive optical tomography (DOT) is a well-established technique for measuring blood volume and oxygen concentration in a variety of tissues. In acousto–photonic imaging (API) a focused ultrasound beam is used to modulate the diffusive photon density wave, resulting in a ''virtual source'' of modulated photon density waves emanating from within the tissue. There are two primary mechanisms for this interaction: index of refraction perturbations and induced particle motion. In this talk, we propose a simple model for the ultrasound-induced motion of a particulate suspension in a fluid. The model accounts for first order oscillatory particle translation and second-order effects such as steady streaming and primary radiation force. Coupled to numerical simulations of the pressure field and of the steady streaming field, this model yields quantitative estimates for the particle motion as a function of space and time, for both continuous wave and pulsed wave insonations. The results are fed into a Monte Carlo simulation of the photon migration/scattering process to yield predictions of the API virtual source strength as a function of acoustic parameters and propagation medium properties. [Work supported by the Center for Subsurface Sensing and Imaging Systems, under the NSF ERC Program (award #EEC-9986821).]

THURSDAY MORNING, 6 DECEMBER 2001 ROOM 221, 9:00 TO 11:45 A.M.

Session 4aSA

Structural Acoustics and Vibration: Numerical Methods in Structural Acoustics I

Scott D. Sommerfeldt, Chair
Department of Physics and Astronomy, Brigham Young University, PN241 ESC PO Box 24673, Provo, Utah 84602-4673

Invited Papers

9:00

4aSA1. On the solution of inverse acoustic scattering problems using shape sensitivity analysis. Gonzalo R. Feijoo, Peter M. Pinsky (Div. of Mech. and Computation, Stanford Univ., Stanford, CA 94305-4040, grfeij@stanford.edu), and Assad A. Oberai (College of Eng., Boston Univ., Boston, MA 02215)

The solution of inverse acoustic scattering problems is considered, i.e., some measures of the pressure field in an acoustic domain are given and we are asked to identify the shape of the object that produces them. The problem is considered in an optimization framework where a cost function is introduced to quantify the error between a guess for the shape of the scatterer and the actual shape. Starting from an initial guess, the algorithm produces a sequence of shapes that converge to a local minima of the functional. At each step, the algorithm requires calculation of the cost function and its derivative with respect to changes in the shape. An efficient method for calculating shape derivatives is introduced. The method is based on the solution of an adjoint problem, which is of the same type as the acoustic scattering problem. The main characteristic of this method is that, irrespective of the number of parameters used to describe the shape of the scatterer, only one adjoint problem is solved for calculating all the derivatives. Numerical examples that demonstrate the efficacy of this approach are presented.

9:30

4aSA2. On the solution of three-dimensional inverse acoustic scattering problems. Charbel Farhat, Radek Tezaur, and Rabia Djellouli (Dept. of Aerosp. Eng. Sci., Univ. of Colorado at Boulder, Boulder, CO 80309-0429)

A computational methodology for retrieving the shape of an obstacle from the knowledge in a limited aperture of its acoustic far-field pattern is presented. This methodology, which features a reconstruction algorithm for the far-field pattern, also distinguishes itself from existing iterative Newton-type procedures by (a) the usage of exact Jacobian matrices for formulating the linearized problems rather than approximate ones (i.e., based on finite differences), (b) a fast numerical procedure for computing these Jacobian

matrices, and (c) a fast and numerically scalable fictitious domain decomposition method for high-frequency acoustic scattering problems. The salient features of this computational methodology are illustrated with several submarine applications. [Work supported by ONR.]

10:00

4aSA3. Local/global methods for structural acoustic radiation and scattering. Donald Bliss and Linda Franzoni (Dept. of Mech. Eng. & Mater. Sci., Duke Univ., Durham, NC 27708-0300)

Solving structural acoustic problems in the mid-frequency range is a challenging task. Computation is expensive due to the large well-coupled nature of the fluid-structure system. Discrete structural elements, e.g., ribs and stiffeners, cause localized discontinuities that require much greater resolution than would be otherwise warranted. Radiation and scattering problems having certain types of constraints exhibit great sensitivity to fine-scale convergence. Our purpose in this research is to devise analytical and numerical ways to recast such problems to be amenable to an efficient and accurate numerical solution. Analytical/Numerical Matching (ANM) decomposes a problem into local and global parts. Analysis methods extract the local discontinuities, replacing the original discrete influence of the inhomogeneity by smooth forces. A smoother computational problem can be solved efficiently, without loss of information. The method has been demonstrated for configurations of increasing complexity, and has become a useful tool for solving realistic problems. The ANM approach has been successfully applied to the modal/FEA structural-acoustics code SONAX, providing a new capability to handle 3-D structural discontinuities. Future research on the application of ANM to other finite element structural-acoustic codes is outlined. An ongoing effort to develop a general theory of structural acoustic homogenization is also summarized.

10:30–10:45 Break

10:45

4aSA4. An efficient algorithm for the simultaneous solution of exterior acoustic problems over a frequency band. Marcus M. Wagner, Peter M. Pinsky (Division of Mech. and Computation, Stanford Univ., Stanford, CA 94305, pinsky@stanford.edu), and Manish Malhotra (Sun Microsystems, 901 San Antonio Rd., Palo Alto, CA 94303, manish.malhotra@sun.com)

A method for solving time-harmonic exterior acoustics problems in a frequency band over selected regions of the computational domain is presented. The partial fields of interest include surfaces enclosing the sound source, as well as distinct points in the near-field of the source. The discretization of the boundary-value problem is based on finite elements. Replacing the infinite domain problem with an equivalent formulation in a bounded domain leads to the incorporation of a Dirichlet-to-Neumann (DtN) map, which accounts for the analytical asymptotic behavior of the solution. By interpreting the discretized DtN operator as a low-rank update of the system matrix, a standard shifted form of the inverse of this matrix, which is incorporated in the expression for the partial fields, is obtained. Due to the complex frequency dependence of this inverse, a direct evaluation is prohibitive. Hence, a matrix-valued Padé approximation of the inverse operator is employed via a two-sided block Lanczos algorithm, which provides a stable and efficient representation of the Padé approximation. A numerical example illustrates the performance and advantages of the outlined method.

Contributed Papers

11:15

4aSA5. Fluid coupling analyses using *in vacuo* modes as basis functions. John B. Fahnline (Appl. Sci. Bldg., Penn State Univ., University Park, PA 16802)

Similar to some of the earliest papers on fluid coupling analyses, the present formulation uses *in vacuo* modes as basis functions, and thus the basic solution technique is well known. However, the inherent difficulties in generating the frequency dependent acoustic impedance and modal coupling matrices have prevented this type of analysis from being applied to practical engineering problems. In the numerical implementation, the matrices are computed using a boundary element analysis and many of the typical limitations are avoided by using a number of recently-developed techniques. First, different structural and acoustic meshes are employed to properly represent the variations in the physical quantities. This is especially important at low frequencies for in-water analyses where the structural wavelength is much smaller than the acoustic. Second, because the acoustic matrices are slowly varying functions of frequency, their individual terms can be interpolated from values calculated at sparse intervals. A preliminary analysis is required to map out the matrices as a function of frequency, but the subsequent frequency response calculations are orders of magnitude faster than if they were performed directly. Several examples are given to demonstrate the validity of the computations, including a cantilever plate, a circular cylinder, and a bass-reflex loudspeaker.

11:30

4aSA6. Analysis of interior and exterior sound fields using iterative boundary element solvers. Martin Ochmann, Laszlo Hetey (Univ. of Applied Sci., TFH-Berlin, Luxemburger Strasse 10, D-13353 Berlin, Germany), and Steffen Marburg (Technische Universitaet Dresden, Institut fuer Festkoerpermechanik, 01062 Dresden, Germany)

The boundary element method (BEM) is an efficient tool for the calculation of interior and exterior sound fields. However, the discretization of the Helmholtz integral equation often leads to large, unsymmetrical, and fully populated systems of linear equations. Such systems can be solved by means of various iterative solvers. Three examples are considered: First, a preconditioned version of the generalized minimum residual method (GMRES) is used for calculating the high-frequency acoustic scattering from cylinderlike structures. Second, the sound field in a passenger compartment is analyzed by using five different iterative solvers: GMRES, the conjugate gradient squared algorithm (CGS), the quasiminimal residual approximation (QMR), the biconjugate gradient stabilized algorithm (BICGSTAB), and the conjugate gradient iteration on the normal equations (CGNR). These methods and a Gaussian elimination are compared with respect to their numerical performance and speed of convergence. Third, scattering and radiation of a nonconvex structure is investigated. This structure—called ''cat's eye''—consists of a sphere, where the positive octant is cut out. The efficiency and accuracy of the above mentioned five iterative methods are studied for the vibrating cat's eye. By means of the preconditioned GMRES, scattering cat's eyes of several thousand boundary elements can be treated on personal computers.

Session 4aSC

Speech Communication: Suprasegmentals, Lexical Effects and Planning (Poster Session)

Bryan W. Gick, Chair
Department of Linguistics, University of British Columbia, 1866 Main Mall, Vancouver, British Columbia V6T 1Z1, Canada

Contributed Papers

All posters will be on display from 8:30 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 10:15 a.m. and contributors of even-numbered papers will be at their posters from 10:15 a.m. to 12:00 noon.

4aSC1. Effect of speaking rate on segment durations of voiceless stops in Korean. Richard Morris and In-sop Kim (107 RRC, Dept. of Commun. Disord., Florida State Univ., Tallahassee, FL 32306-1200)

Studies of the effect of speaking rate on speech segment durations reveal that the VOTs of long lag and prevoiced consonants vary directly with speaking rate, whereas those of short lag consonants remain relatively constant across speaking rates (i.e., Kessinger and Blumstein, 1998; Miller and Volaitis, 1989). Koreans have three voiceless stop consonants at each place of articulation. These stops vary in VOT with values in the long lag and short lag ranges. These phonemes are also marked by differing fundamental frequency patterns. The purpose of this study was to investigate the effects of different speaking rates on the acoustic parameters of the Korean voiceless stop consonants. Sixteen Korean adults (8 male and 8 female) repeated productions of the lax, reinforced, and aspirated forms of /p/, /t/, and /k/ with the vowel /i/. These repetitions were made of the isolated CV syllables and of the syllables in a Korean sentence context. The syllables were analyzed for segment durations of hold, transition and vocalic portions and fundamental frequency contours. The speech was recorded using a DAT recorder and analyzed using a CSL 4300B system. The effect of the speaking rate on the different phonemes will be discussed [R. Kessinger and S. Blumstein, ''Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies,'' J. Phonetics **26**, 117–128 (1998); J. Miller and L. Volaitis, ''Effect of speaking rate on the perceptual structure of a phonetic category,'' Percept. Psychophys. **46**, 505–512 (1989)].

4aSC2. Speaking rates of American and New Zealand varieties of English. Michael Robb (Univ. of Connecticut, Dept. of Commun. Sci., 850 Bolton Rd., Storrs, CT 06269), Yang Chen (Univ. of Wyoming, Dept. of Commun. Disord., Laramie, WY), and Margaret Maclagan (Univ. of Canterbury, Dept. of Speech–Lang. Therapy, Christchurch, New Zealand, m.maclagan@spth.canterbury.ac.nz)

Measures of speaking rate provide an index of speech timing control. The present study sought to evaluate speech timing control in two varieties of English, American English (AE), and New Zealand English (NZE). Measures of speaking rate were calculated for 40 adult AE speakers which were then compared to a group of 40 adult NZE speakers. Results of the analysis identified significantly faster speaking rate and articulation rate among the NZE speakers compared to AE speakers. Vowel changes occurring in both varieties of English over the past decades, as well as contrasting rhythmic properties, are proposed to account for the rate differences. Because temporal differences can be a component of disordered speech and are a major component in certain motor speech disorders and disorders of fluency, normative speaking rate data are often utilized by speech–language pathologists. The present results, viewed from a clinical standpoint, would suggest that the particular variety of English spoken by a client is an important variable when addressing disorders of speech timing.

4aSC3. Effects of speaking rate on temporal patterns of English. Bruce L. Smith (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2299 N. Campus Dr., Evanston, IL 60208 b-smith2@northwestern.edu)

Previous research [B. L. Smith, J. Acoust. Soc. Am. **108**, 2438–2442 (2000)] has indicated that individual subjects show considerable variation in the extent to which they manifest a number of temporal patterns in their speech (e.g., vowel lengthening before voiced obstruents, final-syllable vowel lengthening, etc.). At least in part, some of the variation across speakers was found to be related to the fact that some subjects naturally tend to talk somewhat faster/slower than others. The present study extended these previous findings concerning variations in temporal performance related to rate of speech by having 15 subjects produce several different CVC target stimuli in sentences spoken at both normal and fast rates. The temporal patterns observed as a function of these experimental manipulations of speaking rate generally supported the previous findings related to natural variations in rate of speech across subjects. It was found, for example, that vowel lengthening preceding voiced obstruents tended to be less when subjects spoke at a fast versus a normal rate. In contrast, phrase-final vowel lengthening was typically greater when subjects spoke at a faster rate. Possible reasons for these different vowel-lengthening effects may relate to the temporal domain that influences them.

4aSC4. Exploring underlying pitch targets in English statements. Yi Xu and Ching X. Xu (Dept. of Commun. Sci. and Disord., Northwestern Univ., Evanston, IL 60208)

This study investigates $F0$ contours and their potential association with underlying pitch targets in English statements with and without narrow focus. Eight native speakers of American English read aloud 21 sentences at 2 speaking rates with 7 repetitions. The initial, middle, and final words in these sentences vary in focus status (focus or no focus) and word length (monosyllabic, disyllabic, or trisyllabic). $F0$ curves are extracted using a method that combines automatic vocal cycle detection and manual rectification. Preliminary results show that $F0$ peaks occur earlier when the vowel of the accented syllable is phonologically long than when it is phonologically short. This seems to suggest that a nonfinal pitch accent in an English statement is probably associated with a high static target. Preliminary analyses also show that $F0$ contours of nonaccented syllables are influenced more by the preceding accent than by the following accent. Furthermore, the influence of the accented syllable on the $F0$ of the fol-

lowing (but not the preceding) nonaccented syllable is found to reduce over time. These patterns seem to suggest that inter-accent $F0$ contours in English probably also result from certain underlying pitch targets rather than from linear or nonlinear interpolation between adjacent pitch accents.

4aSC5. Contrastive focus in Beijing Mandarin. Yiya Chen (Linguist. Dept., SUNY-Stony Brook, Stony Brook, NY 11794-4367)

Contrastive focus in Beijing Mandarin. This study investigated the acoustic realization of contrastive focus in Beijing Mandarin. Here a contrastive focus is defined as a mechanism to signal contrast or make corrections in discourse. Two questions were asked. First, what are the acoustic cues employed to convey contrastive focus? Second, what are the linguistic constraints on the manifestation of prosodic cues? Four different renditions of each test sentence were elicited from the subjects with contrastive focus on different words to convey different pragmatic meanings. Preliminary results suggest: (1) All three acoustic cues under investigation $F0$, overall intensity, and duration are employed, with different degrees of consistency, to convey contrastive focus; (2) the manifestation of these cues also appear to be context sensitive. The implications of these phonetic facts for possible phonological organization of contrastive focus in Beijing Mandarin will be discussed. [Work supported by NSF Grant No. SBR 9600930 to Marie Huffman.]

4aSC6. Toward a perception based model of the production of prosody. Grzegorz Dogil and Bernd Moebius (Inst. of Natural Lang. Processing, Univ. of Stuttgart, Azenbergstr. 12, D-70174 Stuttgart, Germany)

The speech production model proposed by Guenther, Perkell, and coworker's [Guenther *et al.*, Psych. Rev. **15** (1998); Perkell *et al.*, J. Phon. **28** (2000)] is generalized from the predominantly segmental perspective to a new theory of the production of prosody. It is posited that the only invariant targets of the production of prosodic events are perceptual targets, characterized as multidimensional regions in the perceptual space. Prosodic gestures are planned as trajectories that reach and traverse the perceptual target regions. Successfully executed gestures produce acoustic realizations of perceptually relevant prosodic events. The prosodic interpretation of the speech production model is structured around a hierarchy of prosodic domains: discourse structure, information structure, and accentual structure. It is assumed that the linguistically relevant and phonologically distinctive functions of prosodic features are represented by internal models. Once acquired, the phonemic settings pertinent to these models are stable and resistant to change, and they do not rely on continuous auditory feedback. We further suggest that the mechanisms involved in prosody control, may not differ categorically from those that control segmental speech production.

4aSC7. Preliminary studies on respiratory activity in speech. Peter Ladefoged (Dept. of Linguist., UCLA, Los Angeles, CA 90095-1543, oldfogey@ucla.edu) and Gerald Loeb (Biomed. Eng., USC, Los Angeles, CA 90089-1112)

Work at Edinburgh University in the 1950s, summarized by Ladefoged [*Three Area of Experimental Phonetics* (Oxford University Press, Oxford, 1967)] claimed that (1) the mean power for speech was provided by actions of the respiratory muscles. The external intercostals check the outgoing airflow when speaking with high lung volumes, and the internal intercostals, rectus abdominis, and other muscles become more active as lung volume decreases. (2) Stressed syllables are produced by increases of respiratory power above the mean, usually by further activity of the internal intercostals. Hixon and Weismer (1995) pointed out technical limitations of these early studies, but wrongly indicated that they were of little value. These studies have been replicated with modern methods to measure intramuscular EMG and airflow and pressure. The two subjects studied to date follow the pattern previously described. In addition the improved techniques indicate that speakers exercise further control by bringing in stabilizing antagonistic muscles. These results are consistent with the general hypothesis that the central nervous system is able to control subglottal pressure by using various combinations of passive elastic and active muscle forces, thereby achieving motor equivalence and acoustic invariance over a wide range of lung volumes.

4aSC8. Stress effects on voicing and frication in voiced obstruents in north-central Spanish. Carolina Gonzalez (USC, Linguist. Dept., GFS 301, 3601 Watt Way, Los Angeles, CA 90089-1693)

In Spanish /b, d, g/ are usually realized as voiced approximants in all sylllabic contexts unless following a stop—a phenomenon called spirantization. However, in North-Central Peninsular Spanish (NCS) voiceless fricatives rather than voiced approximants are produced. In the majority of cases this occurs in stressed syllables. This study examines whether stress is a factor in the likelihood of frication and devoicing of coda /b, d, g/ in this dialect. Nine native speakers of NCS were recorded producing nonce words with /b, d, g/ in coda position in both stressed and unstressed syllables. Measurements were made of vowel and consonant duration, presence and absence of frication and voicing, and voicing duration. The results indicate that frication is more likely in stressed syllables than in unstressed syllables. This suggests that in stressed syllables a higher subglottal pressure produces higher airflow across the glottis, favoring frication. In turn, frication inhibits voicing due to conflicting aerodynamic requirements between the two. It is concluded that stress is a factor in spirantization and that it may indirectly affect the voicing properties of /b, d, g/. [Work supported by Del Amo and the Basque Government.]

4aSC9. Structural constraints in the perception of English stop-sonorant clusters. Elliott Moreton (Dept. of Cognit. Sci., Johns Hopkins Univ., Baltimore, MD 21218)

Native-language phonemes combined in a non-native way can be misperceived so as to conform to native phonotactics; e.g., English listeners are biased to hear syllable-initial [tr] rather than the illegal [tl] (Massaro and Cohen, 1983; Pitt, 1998). What sort of linguistic knowledge causes phonotactic perceptual bias? Two classes of models were compared: unit models, which attribute bias to the listeners differing experience of each cluster (such as their different frequencies); and structure models, which use abstract phonological generalizations (such as a ban on [coronal][coronal] sequences). Listeners ($N=16$ in each experiment) judged synthetic 6×6 arrays of stop-sonorant clusters in which both consonants were ambiguous. The effect of the stop judgment on the log odds ratio of the sonorant judgment was assessed separately for each stimulus token to provide a stimulus-independent measure of bias. Experiment 1 compared perceptual bias against the onsets [bw] and [dl], which violate different structural constraints but are both of zero frequency. Experiment 2 compared bias against [dl] in CCV and VCCV contexts, to investigate the interaction of syllabification with segmentism and to rule out a compensation-for-coarticulation account of Experiment 1. Results of both experiments favor the structure models. [Work supported by NSF IGERT.]

4aSC10. The influences of phonotactic constraints of words on speech production. Patricia Ward and Jan Charles-Luce (Univ. at Buffalo, 122 Cary Hall, 3435 Main St., Buffalo, NY 14214)

It is known that words are stored in the mental lexicon with certain lexical and phonological properties and that these properties have consequences for the perception and production of words. Phonotactic probability is a phonological property of words. It is defined by segment position frequency and bi-phone co-occurrence frequency. A word can have a high probability pattern or a low probability pattern. In this study, the effects of the phonotactic probabilities of words on speech production have been examined in children with typical speech and language skills and children with phonological disorders. Typical children between the ages of 4–6,

7–8, and 9–11, children with phonological disorders, ages 4–11, and typical adults were studied. Twelve CVC words and nonwords, six high probability and six low probability, were used as stimuli. Each subject produced the stimulus items for audio-recording and later waveform analysis. Acoustic measurements, specifically duration, differed based upon the phonotactic probability of the stimulus item. Differences in the acoustic measurements of these productions by children of varying ages and language abilities reflect the influences of phonological and lexical variables of words stored in the mental lexicon. Results will be discussed with respect to developmental implications and cognitive skills.

4aSC11. Word perception in time-compressed speech. Hugo Quené[a] and Esther Janse (Utrecht Inst. of Linguist, Utrecht Univ., Trans 10, NL-3512 JK Utrecht, The Netherlands, hquene@indiana.edu)

Certain applications, e.g., voicemail playback or text read-out, time-compress their output speech, usually in a linear fashion. This paper compares spoken-word perception, as measured by phoneme monitoring, in natural fast-rate speech and in normal-rate speech that is linearly time-compressed to the same fast rate. Perception is predicted to be more difficult in the natural-fast condition, in spite of its naturalness, due to its reduced articulation and intelligibility. Results show that listeners indeed reacted considerably faster in the time-compressed condition, as predicted. Nevertheless, the time-compressed condition suffers from inappropriate timing patterns, since fast-rate speech timing patterns differ from normal-rate speech on which it is based. In a second experiment, it is investigated whether perception can be improved by providing listeners with the appropriate timing patterns. To this end, the timing patterns (segmental durations) of time-compressed speech are manipulated to equal those of natural-fast speech. Hence the effects of segmental intelligibility and of temporal patterns can be separated. In summary, results obtained so far show that spoken-word perception in natural-fast speech is hampered by speakers reduced articulation. In time-compressed speech, the perceptual advantages of improved segmental intelligibility are apparently more important than the abnormality of hearing very fast but fully articulated speech. [a] Hugo Quené is currently at Linguistics Department, Indiana University, Bloomington, Indiana.

4aSC12. Positional and planning effects on the durational structure of repetition strings. Alan Bell and Cynthia Girand (Dept. of Linguist., Univ. of Colorado, Boulder, CO 80309, abell@psych.colorado.edu)

Gaining time to resolve some difficulty in the production of upcoming speech is the primary function of unplanned repetitions. Understanding their duration structure is thus crucial to modeling their production, which surely differs greatly from fluent phrases. When the duration structure of the entire repetition string of unplanned repetitions is examined, strong global dependencies are found. Repetition strings are words/phrases repeated once or more, together with silent and filled pauses optionally occurring next to them. The main effects are that durations of first and second string items, whether repeated words or pauses, are positively correlated with the duration of the rest of the string; items, whether words or pauses, are shorter as they occur later in the string; and strings that begin with a pause average longer than strings that do not. The study is based on an analysis of 503 disfluent repetitions taken from the ICSI phonetically transcribed sample of the Switchboard conversation corpus, extensively checked and recoded. The results help explain local durational dependencies of the repeated words [Bell and Girand, LabPhon 7 (2000)] and imply that the articulation of repetitions is influenced from the onset by the nature and degree of difficulty they address.

4aSC13. Frequency and predictability effects on the duration of content words in conversation. Michelle Gregory (Dept. of Linguist. and Cognit. Sci., Brown Univ., Providence RI 02912, michelle_gregory@brown.edu), Alan Bell, Daniel Jurafsky (Dept. of Linguist., Univ. of Colorado, Boulder, CO 80309), and William Raymond (Dept. of Linguist., Ohio State Univ., Columbus, OH 43210)

This study carefully examines the proposal that word forms in conversation are shorter when they have higher probability, where this includes all factors of their context (Jurafsky *et al.*, *Frequency and the Emergence of Linguistic Structure*, edited by Bybee and Hopper, 2001, pp. 229–254). Higher word frequencies have long been known to be linked to shorter durations. But it has been hard to confirm a direct relationship between the two. One confound is with predictability or conditional probability, which also affects surface form and is closely related to frequency. Another difficulty is controlling factors related both to duration and contextual probability. Linear regression (and sampling to avoid violating independence assumptions) was used to control for phonological form, speech rate, prosodic prominence, previous uses, and neighboring disfluencies. Both greater word frequencies and greater conditional probabilities given the following word have a strong shortening effect on durations of content words. The effect of other probability measures considered did not reach significance. The study is based on a sample of about 1000 content words from a portion of the Switchboard corpus, phonetically transcribed at ICSI, Berkeley, and coded for intonation by P. Taylor, M. Ostendorf, and S. Shattuck-Hufnagel. [Work supported by NSF.]

4aSC14. Frequency and category factors in the reduction and assimilation of function words. Dara Kanwischer, Bryan Gick, Ian Wilson (Univ. of British Columbia, Dept. of Linguist, E270-1866 Main Mall, Vancouver, BC V6T 1Z1, Canada, gick@interchange.ubc.ca), and Rushen Shi (Univ. of British Columbia, Vancouver, BC V6T 1Z3, Canada)

It has been observed that there are phonetic and phonological differences between function words and content words. However, it is unclear whether these differences are actually the result of the distinct syntactic categories, or of other secondary effects such as frequency or stress. A production experiment was conducted to distinguish between the factors of category and frequency while controlling for stress. EPG and acoustic data were collected for five subjects. The effects of the two factors were examined within several processes of reduction and assimilation, including final-t/d dropping, initial-/h/ dropping, and palatalization. Preliminary results suggest that both word frequency and category are important factors in these processes. The consonant-dropping processes seem to show a strong sensitivity to category, but also show within-category effects of frequency. Palatalization appears to be a more gradient phonetic process, with results thus far suggesting greater sensitivity to frequency than to category. The fact that the same processes were affected by both factors suggests that at least some of the phonetic and phonological properties traditionally assigned to function words as a category may be wholly or partly the result of high word frequency. [Research supported by NSERC.]

4aSC15. Representational specificity of lexical form in the perception of spoken words. Conor T. McLennan, Paul A. Luce, and Jan Charles-Luce (Lang. Percept. Lab., Univ. at Buffalo, 245 Park Hall, Buffalo, NY 14260)

The specificity of form-based lexical representations of spoken words was examined. Using a repetition-priming paradigm, we attempted to determine (1) if flapped intervocalic alveolar stops in American English are mapped onto underlying representations of /t/, /d/, or both or (2) if flaps have an independent representational status. Participants shadowed or made lexical decisions to spoken words in two blocks of trials. Stimuli in the first block served as primes and those in the second block as targets. Primes and targets consisted of flapped and carefully articulated bisyllabic

words. We measured reaction times to target words in the second block as a function of prime type. The results provide evidence for both underlying and surface perceptual representations.

4aSC16. Effects of subphonemic variation depend on lexical competitor environment. Petra van Alphen (Max Planck Inst. for Psycholinguist., Wundtlaan 1, 6525 XD Nijmegen, The Netherlands)

A cross-modal identity priming experiment examined whether subphonemic variation (length of prevoicing in syllable-initial stops in Dutch) influences lexical access. Word and nonword targets began with voiced plosives. They were chosen so that half of them became words when the initial voiced plosive was replaced by its voiceless counterpart, while the other half changed into nonwords. The visual targets were preceded by auditory word or nonword primes which were either phonologically unrelated or identical to the target. The initial plosives of the identity primes had 0, 6, or 12 periods of prevoicing, but all were identified as voiced. This VOT manipulation had no effect for word targets: lexical decisions were faster after all three identity primes than after the unrelated primes. However, reaction times to nonword targets with word competitors (e.g., BRINS, with the competitor PRINS ''prince'') were slower when preceded by identity primes without prevoicing than when preceded by primes starting with 6 or 12 periods of prevoicing. This suggests that voiceless competitors are activated by words starting with voiced plosives without prevoicing. Subphonemic variation appears to influence the degree of lexical activation, but this effect depends on the lexical competitor environment.

4aSC17. Phonological neighborhoods and phonetic similarity in spoken word recognition. Kiyoko Yoneyama (Dept. of Linguist., Ohio State Univ., Columbus, OH 43210)

English neighborhood literature has demonstrated that neighborhood density affects the auditory word recognition. However, an unresolved question is exactly how neighborhood density should be calculated. In this paper the definition of a lexical neighborhood is explored in Japanese. Data for the analyses were collected from Japanese neighborhood experiments using the same 700 test words and a lexicon that consisted of only nouns from the NTT psycholinguistic database [Amano and Konodo, 1999]. Three different neighborhood calculations were used to analyze the data. The first calculation was based on Greenberg–Jenkins phoneme substitution, deletion, and insertion rules. The second calculation included prosodic information as another dimension in the neighborhood calculation in order to reflect the finding that prosodic information has a vital role in Japanese word recognition. The third calculation was based on the auditory properties of the words in the lexicon. Neighborhood density was measured by comparing the similarity of cochleagrams of the 66 000 audio files. The results of the analyses demonstrated that phonological similarity within the lexicon seems to be calculated based on higher-level abstract representations rather than on a lower-level acoustic-auditory representation in any of the experiments. The implications from the results to the current word recognition theories will also be discussed. [Work supported by NIH.]

4aSC18. Talker and speaking rate variation affect lexical neighborhoods. James R. Sawusch, Liza K. Zimack, Kathleen M. Measer (Psych. Dept., Univ. at Buffalo, Park Hall, Buffalo, NY 14051), Rochelle S. Newman (Psych. Dept., Univ. of Iowa, 11 Seashore Hall E, Iowa City, IA 52242), and Paul A. Luce (Psych. Dept., Univ. at Buffalo, Park Hall, Buffalo, NY 14051)

Words that are similar to a syllable can influence listeners perception of phonemes in the syllable. This influence of lexical neighborhood is part of how the words of a language influence speech perception. In previous studies (ASA, June 2000 and Dec 2000), this effect was shown to vary with talker voice, indicating that the similarity of words to nonsense syllables (lexical neighborhood) varies with talker. In a new series of studies, token variation in the nonsense syllables was introduced. Listeners identified the syllable initial stop in bowth-powth and powse-powse series. In these series, bowth and powse are similar to more real words. One group of listeners heard these series as spoken by four different talkers. The tokens of the different talkers were randomly intermixed. A second group heard tokens from one talker at three different speaking rates: fast, normal, and slow. Again, these were randomly intermixed. To the extent that token variation alters the influence of lexical neighborhood, factors other than simple acoustic–phonetic similarity must play a role in determining lexical neighborhood. The implications of these data for models of word recognition will be discussed. [Work supported by NIDCD Grant No. R01DC00219 to SUNY at Buffalo.]

THURSDAY MORNING, 6 DECEMBER 2001 ROOM 223, 8:30 TO 11:45 A.M.

Session 4aSP

Signal Processing in Acoustics: General Topics in Signal Processing I

Leon H. Sibul, Chair

Applied Research Laboratory, Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804-0030

Contributed Papers

8:30

4aSP1. Application of extended Kalman filtering to tracking of airborne high-speed broadband acoustic sources. William G. Frazier, Chad M. Williams, Jay E. Williams (Miltec, Inc., NCPA, Coliseum Dr., University, MS 38677), and Kenneth E. Gilbert (Univ. of Mississippi, Natl. Ctr. for Physical Acoust., Coliseum Dr., University, MS 38677)

A technique for tracking airborne high-speed (subsonic) broadband acoustic sources is presented. This technique is based upon the familiar extended Kalman filter algorithm used in navigation, radar-based tracking, and feedback control applications. In this application, acoustic sensor-to-source bearing measurements from two locations are used to estimate target position, speed, and heading in near real time. In particular, two strategies for compensating the variable delay in obtaining bearing information caused by sound propagation are developed and presented. One of these is based upon a simple linear predictor. While adequate for sources that are not too distant from the acoustic sensors, it is shown to perform inadequately in some longer-range situations. A second approach, based upon reformulating the Kalman filter to work in conjunction with a generalized triangulation algorithm, is shown to be more robust. Advantageous features of employing a Kalman filter approach to tracking problems include the ability to systematically compensate for measurement noise and to accommodate target maneuverability constraints. Application

of the technique to data obtained from a highly maneuverable low-flying commercial aircraft are presented and compared to flight path data obtained via the Global Positioning System.

8:45

4aSP2. Iterative model-based automatic target recognition for acoustic landmine detection. Ning Xiang and James Sabatier (Univ. of Mississippi, Natl. Ctr. for Physical Acoust., 1 Coliseum Dr., University, MS 38677)

A model has been developed to allow the scanned data obtained using a laser Doppler vibrometer-based acoustic-to-seismic landmine detection system to be analyzed without operator interaction. The ground vibration data from the LDV are pre-processed to form images in a 2-D data format. A parametric model was established to describe the amplified magnitude velocity phenomena induced by buried landmines. This model incorporates amplitude, size, position, and background amplitude parameters into an automatic analysis process. An iterative regression approach is described which can perform the automatic recognition of landmines. The estimated parameters, such as the amplitude relative to the background, the size, and the shape of a target, are used to make the decision regarding the presence of a mine. Once a positive decision is made, the estimated position parameters are used to localize the target location. [Work supported by the United States Army Communications-Electronics Command Night Vision and Electronic Sensors Directorate.]

9:00

4aSP3. Measured performance of an endfire superdirective line array in littoral water. Paul C. Hines, Daniel L. Hutt, and Victor Young (Defence Res. Establishment Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada)

Superdirective line arrays can provide high gains whenever the interelement spacing is much less than half a wavelength. In practice these arrays have fallen short for two reasons: First, performance degrades appreciably in the presence of uncorrelated system noise. Second, the optimum array weights are specific to the form of the noise field and the choice of these weights is often sensitive to deviations from that noise field. Furthermore, since the process involves computing signal differences between hydrophones, the signal to noise ratio (SNR) degrades as the steering angle approaches broadside. In spite of these drawbacks, a superdirective array can provide substantial improvement over a conventional array in specific instances. For example, the present application requires an endfire array of limited extent (1 m in length) and wide bandwidth (1–5 kHz) making a conventional array impractical. Moreover, improvements in array hardware have substantially reduced uncorrelated system noise, and increased computer speed makes it reasonable to process data using several weighting schemes. In this paper, the superdirective array will be described and its measured performance will be compared to theoretical predictions.

9:15

4aSP4. Measurement of underwater sound intensity and determination of source bearing angle. Daniel Hutt, Paul Hines, and Victor Young (Defence Res. Establishment Atlantic, 9 Grove St., Dartmouth, NS B2Y 327, Canada)

Measurements of the underwater acoustic pressure and pressure gradient can be used to calculate the components of the acoustic intensity vector. The intensity vector points in the direction of net power flow and can therefore be used to find the direction or bearing angle to an acoustic source. Use of intensity measurements for determination of bearing is of interest because high-angular resolution can be achieved using a relatively small array with physical dimensions less than 1/4 of the wavelength. The usefulness of this technique was investigated in a series of experiments on the Scotian Shelf in 70 m of water. The intensity sensor is an array of six hydrophones arranged as three orthogonal dipoles. The hydrophone spacing is 11.4 cm with a frequency range of 500–4000 Hz. For signal-to-noise ratios greater than 10 dB, bearing angles to a source were determined with a standard deviation of less than 0.1 deg. However, offsets of several degrees from the correct bearing angle were often present. The offset error is believed to be due to acoustical scatter from supporting structures. The effect of noise, gain error, and phase error on intensity measurements is discussed.

9:30

4aSP5. Resolution of left–right ambiguity in towed array sonars. Katerina Kaouri and David J. Allwright (Ctr. for Industrial and Appl. Math., Mathematical Inst., Oxford Univ., 24-29 St. Giles', Oxford OX1 3LB, UK, kaouri@maths.ox.ac.uk)

A method is proposed for resolving the left–right ambiguity in passive sonar systems, a problem that arises in source localization when the array is straight. In practice, the array is *not* straight and a statistical analysis within the Neyman–Pearson framework is developed for a monochromatic signal in the presence of noise, assuming that the exact array shape is known. For any given array shape, an expression for the probability of correct resolution (PCR) is derived as a function of two parameters: the SNR and an array-lateral-displacement parameter. The first parameter measures the strength of the signal relative to the noise and the second the curvature of the array relative to the acoustic wavelength. The PCR is calculated numerically for a variety of array shapes of practical importance. The model results are found to agree with intuition; the PCR is 0.5 when the array is straight and is increasing as the signal is becoming louder and the array more curved. It is explained why the method is useful in practice and the effects of correlation between beams are discussed. [Research partially supported by Thomson Marconi Sonar Ltd.]

9:45

4aSP6. Characterization of the left/right bearing ambiguities for passive line arrays. Frank A. Boyle and David E. Grant (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029)

The symmetry of a straight horizontal line array produces a left/right bearing ambiguity that makes it impossible to determine on which side of the array a source is located. A simple technique for resolving left/right ambiguities is to break the symmetry by curving the array. Under certain common conditions, however, significantly curved arrays still possess left/right ambiguities that depend on the source's range and the beamformer's focal length. These ambiguities can be modeled, given a source and array configuration, to produce a prediction of source localization capability. Comparisons with actual data were generally consistent with modeled results in the cases considered. The presentation will include a description of how the ambiguities are modeled, and a comparison between modeled and experimental results. Considerations on how curved arrays can be configured to minimize ambiguities will also be discussed.

10:00–10:15 Break

10:15

4aSP7. Deconvolution of chirp sidescan sonar data. Juliette W. Ioup (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148), Maria T. Kalcic, L. Dale Bibee (Naval Res. Lab., Stennis Space Ctr., MS 39529), Edit J. Kaminsky, and George E. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

The first return of a sidescan sonar chirp contains information about the bottom and immediate subbottom character and roughness. Measured real data from a chirp sidescan sonar are basebanded, producing complex signals. These signals are deconvolved using an ideal transmitted source signal, a measured source signal, and, for cross-correlated data, the autocorrelation of the source. Four deconvolution techniques are used: (1) Fourier division in the frequency domain; (2) a least squares technique in the time domain; (3) the reblurring iterative deconvolution method of Kawata and Ichioka (which by definition uses the autocorrelation of the source); and (4) an alternative always-convergent modification of van Cittert iterative deconvolution. The deconvolved signals are compared to the

cross-correlation of the source with the received signal (matched filter) without deconvolution. Results are discussed in relation to the known bottom characteristics. [Research supported in part by an NRL/ASEE Summer Faculty Fellowship and ONR.]

10:30

4aSP8. Wavelet denoising of sidescan sonar images. Juliette W. Ioup (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148), Marlin L. Gendron, Brian S. Bourgeois (Naval Res. Lab., Stennis Space Ctr., MS 39529), and George E. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

One important application of wavelet transforms is for noise removal or denoising. The effectiveness of this technique is influenced by the choice of wavelet used, the decomposition level, and the threshold (both amplitude and type). Thirty different wavelets, several allowable decomposition levels, and a range of appropriate thresholds are tested. Preliminary results will be presented of wavelet denoising applied to 2-D acoustic backscatter imagery from a sidescan sonar in an attempt to improve the detection of bottom features. Comparisons with Fourier based filtering are also discussed. [Research supported in part by an NRL/ASEE Summer Faculty Fellowship and ONR.]

10:45

4aSP9. Automatic identification of sound source direction based on neural networks. Andrzej Czyzewski and Rafal Krolikowski (Sound and Vision Eng. Dept., Tech. Univ. of Gdansk, Narutowicza 11/12, 80-952 Gdansk, Poland)

Automatic identification of sound sources direction is still an unsolved problem in contemporary teleconferencing systems. Speech signals coming from various directions not only interfere with the target signal but also can obscure it. People have difficulty in understanding speech with background noise, high reverberation, and/or with many concurrent speakers. Source identification system should allow tracking a target speaker automatically without much delay in order to avoid picking up concurrent speakers by the same microphone channel. In literature one can find two approaches to this problem. One of them is a classical approach to this problem based on delay-summation algorithms, superdirective arrays and adaptive algorithms, and nonlinear frequency domain microphone array beamformers, etc. The second solution to this problem was proposed among others by the authors in their previous studies, namely, they applied Artificial Neural Networks (ANNs) for the purpose of the automatic sound source localization. In this paper a method for automatic detection of sound source was studied. Both standard feed-forward ANNs and Recurrent Neural Networks were employed for that purpose. Comparison of the results obtained is given. Conclusions are also derived. [Research sponsored by the Committee for Scientific Research, Warsaw, Poland, Grant No. 8 T11D 00218.]

11:00

4aSP10. Using psycho-acoustics frequency masking to reduce distortion in a parametric array speaker. Yew Hin Liew, Kelvin Lee, Furi Andi Karnapi, and Woon Seng Gan (DSP Lab, School of Elec. and Electron. Eng., S2-b4a-03, Nanyang Technolog. Univ., Nanyang Ave., Singapore 639798)

Using a parametric array of ultrasonic transducers, directional sound can be generated from inaudible ultrasonic as a result of nonlinear interaction of ultrasonic wave in air. A finite amplitude ultrasonic wave that can be amplitude modulated by any audio signal is radiated from the array [Yoneyama *et al.*, J. Acoust. Soc. Am. **73**, 1532–1536 (1983)]. As a result, an audio signal is produced because of the self-demodulation effect of the modulated sound wave. This, however, generates a large distortion that is impractical. Another method [T. Kite *et al.*, J. Acoust. Soc. Am. **103**, 2871 (1998)] uses a preprocessing technique, which utilizes double integration of the square-rooted input. The resulted distortion is greatly reduced but at the expense of very wide bandwidth requirement due to the square-root operation. Various techniques to reduce the distortion by preprocessing have been discussed in previous literature. In this paper, we propose a psycho-acoustic masking scheme, coupled with a suitable modulation that will provide further reduction in distortion. The psycho-acoustic masking focus on the distortion occurred within the critical bands. Simulation and subjective testing have been done to verify the system.

11:15

4aSP11. Loudspeaker linearization technique for parametric array in air. Kelvin Lee, Yew-Hin Liew, Furi Andi Karnapi, and Woon-Seng Gan (DSP Lab, School of Elec. and Electron. Eng., S2-B4a-03, Nanyang Technolog. Univ., Singapore)

The use of a parametric array in air as a directional audio loudspeaker has been reported in previous literature. The generation of highly directional audible sound is made possible by the self-demodulation of a finite amplitude ultrasonic carrier which is amplitude-modulated with the audio signal. This self-demodulation process, which is a result of the airs nonlinear response, has undesirable side effects on the audible sound as it is subject to nonlinear distortion. The most widely accepted preprocessing technique for reducing this distortion is based on Bertays far field solution and involves using a square root operation which requires the use of an ultrasonic transducer with infinite bandwidth. We propose a novel approach to model the airs nonlinear response *a priori* using a nonlinear system identification process commonly used in conventional loudspeaker linearization. Inverse modeling of the identified nonlinear system is then performed online and used to predistort the audio signal before it is modulated onto the ultrasonic carrier. A modified Volterra filter structure, known as Multi-memory Decomposition (MMD) and first proposed by Frank *et al.*, is used for the whole linearization process because of its simplicity and efficiency.

11:30

4aSP12. Method to enhance the low-frequency perception from a parametric array loudspeaker. Furi Andi Karnapi, Yew Hin Liew, Kelvin Lee, and Woon Seng Gan (DSP Lab, School of Elec. and Electron. Eng., S2-B4a-03, Nanyang Technolog. Univ., Singapore)

Human auditory system does not perceive the sound signal of all frequencies with equal loudness. Based on equal loudness curves, it is known that a low-frequency signal needs to be produced with a higher-power level to have the same loudness as the middle frequency part. There are two ways to overcome this problem; either boosting the power of the low-frequency part or utilizing the psychoacoustics effect called the missing fundamental. There are many works which have been done utilizing psychoacoustics to enhance the low-frequency perception of audio signal. These works are implemented to normal audio systems. Normal audio systems are comprised of a loudspeaker as the audio reproduction end. It is proposed to implement the low-frequency enhancement to a parametric array loudspeaker. The use of a parametric array loudspeaker to generate highly directional audible signal has been reported in several literatures. However, this audible signal lacks low-frequency content. One of the reasons is the relatively low-power level produced by the existing parametric array. By utilizing the nonlinearity property of the air itself, it is proposed to psychoacoustically enhance the perception of the low-frequency part.

Session 4aUW

Underwater Acoustics: Reverberation in Shallow Water I

Kevin D. LePage, Chair
SACLANTCEN, CMR 426, APO, AE 09613-5000

Chair's Introduction—7:55

Invited Papers

8:00

4aUW1. Some shallow water reverberation highlights and bottom parameter extractions in the 80–4000 Hz region from the boundary characterization and geoclutter experiments. John R. Preston (Appl. Res. Lab., The Penn State Univ., State College, PA 16804) and Dale D. Ellis (Defence Res. Establishment Atlantic, Dartmouth, NS B2Y 3Z7, Canada)

Together with SACLANTCEN, the authors recently participated in two trials to measure shallow water bottom reverberation near Sicily and the STRATAFORM off New Jersey. Sources were monostatic SUS charges and coherent pulses. The receivers were horizontal arrays. Data were analyzed in bands from 80 to 4000 Hz. Highlights of the reverberant returns are discussed. The STRATAFORM is known to have benign surface morphology but contains many buried river channels. One author was also involved in the 2001 Geoclutter experiment conducted in STRATAFORM using coherent bistatic sources near 400 Hz. Differences in reverberation characteristics between SUS and Geoclutter results are discussed. Another objective of these reverberation experiments was to quickly invert for scattering and geoacoustic parameters. A simulated annealing (SA) algorithm was used together with the Generic Sonar Model (NUWC). After automatically adjusting bottom loss and scattering strength, good agreement is achieved between the diffuse reverberation data and model predictions in relatively flat areas. Model/data differences are generally correlated with bottom scattering features. Since reverberation typically lasts 10–20 s or more, extracted parameters apply over wide areas. Local bottom loss and backscattering measurements were made by Holland in these areas. A comparison with Holland's results is given. [Work supported by ONR Code 32, Grant No. N00014-97-1-1034.]

8:25

4aUW2. The Geoclutter experiment: Remotely imaging sub-bottom geomorphology in shallow water. Nicholas C. Makris, Purnima Ratilal, Yisan Lai, and Edward Scheer (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

Preliminary results of the Geoclutter experiment of April–May 2001 are presented. This bistatic reverberation experiment was conducted at the New Jersey Strataform site in continental shelf waters. Roughly 3000 waveforms were transmitted from vertical source arrays in the 390–440 Hz band and received by a horizontally towed array. On average roughly 10–100 discrete clutter events were registered per transmission for a total of at least 30 000 geoclutter events that could be confused with a discrete target over the period of the experiment. Bathymetry throughout the New Jersey Strataform area is extremely benign with few discrete features. The vast majority of clutter returns appeared in level areas where absolutely no bathymetric features are present. The clutter events often registered well with buried river channel networks previously characterized in the Geoclutter program. Extensive probing of river-channel-infested areas revealed that the number and level of clutter events is highly dependent on bistatic orientation. The registrations suggest that bistatic aspects that produce ''glints'' from large projected areas along river channel axes yield the strongest and most frequent targetlike returns. These results are consistent with those predicted by a unified model for reverberation and submerged-target scattering [J. Acoust. Soc. Am. **109**, 909–941 (2001)].

8:50

4aUW3. Use of model predictions to interpret shallow water directional reverberation data. Dale D. Ellis (Defence Res. Establishment Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada) and John R. Preston (Appl. Res. Lab., The Pennsylvania State Univ., State College, PA 16804)

Most shallow water environments are variable over relatively short ranges. Low-frequency reverberation is typically dominated by bottom scattering, and therefore has considerable dependence on range and azimuth. Directional receivers, such as towed line arrays, are an important tool for measuring this azimuthal dependence. Not many reverberation models are capable of handling these beam patterns, but two which have proved useful are the Generic Sonar Model developed at NUWC and the OGOPOGO normal-mode model developed at DREA. By adjusting the bottom scattering strength and sometimes the bottom loss, good agreement has been achieved between the reverberation data and model predictions in relatively flat areas. Differences between the ''flat-bottom'' model predictions and data are generally correlated with bottom scattering features. The talk will focus on comparisons of model predictions

with towed array data obtained in joint trials with the SACLANT Undersea Research Center. The comparisons can be used to extract estimates of average bottom loss and scattering parameters, and to produce crude scattering maps of the bottom. These maps of scattering anomalies, showing the differences between the measurements and the smoothed flat-bottom model predictions, indicate the challenges for range-dependent and more sophisticated models.

9:15

4aUW4. Predicting mono and bistatic reverberation for shallow water waveguides with detailed environmental descriptions. Kevin D. LePage (SACLANT Undersea Res. Ctr., Viale San Bartolomeo 400, 19138 La Spezia, Italy, lepage@saclantc.nato.int)

Over the past several years a suite of mono and bistatic scattering and reverberation tools have been developed at SACLANTCEN for the purpose of predicting reverberation in range dependent shallow water waveguides where high-resolution environmental information is available. Here the theoretical underpinnings of these models is described with particular attention being paid to the following issues: (1) the importance of coherent effects of propagation in reverberation; (2) the effects of range dependence of bottom properties, bathymetry, and sound speed profile on propagation and scatterer excitation; (3) the importance of the spatial scales and distributional properties of the scatterers; and (4) the significance of coherent effects in backscatter from features. Examples of predictions are compared to experimental measurements from the Malta Plateau where supporting high-resolution environmental descriptions are available.

9:40

4aUW5. Technical aspects of the comprehensive acoustic simulation system (CASS) reverberation model and comparisons with measured data. Henry Weinberg (Anteon Corp., 23 Colonial Dr., Waterford, CT 06385) and Ruth Keenan (SAIC, P.O. Box 658, 13 Steeple St., Mashpee Commons, Mashpee, MA 02649, rkeenan@capecod.net)

The Comprehensive Acoustic Simulation System (CASS) has recently received interim acceptance into the United States Navy's Atmospheric and Oceanographic Master Library (OAML). This paper focuses on the application of the CASS to computing reverberation in shallow water. The first topic reviews the numerical approach. This is based on the usual assumption that reverberation received at a specific time can be found by integrating returns from ensonified scattering areas on the ocean boundaries and from ensonified volumes within the ocean proper. The same approach is used for monostatic and bistatic modes of operation. CASS contains various methods for integrating over bearing and frequency. The second topic compares computed results with torpedo data measured off the waters of southern California. This data confirms the requirement for including range-dependent effects in shallow-water reverberation modeling. Fortunately, the measured reverberation was accompanied by an excellent survey of bottom properties. Since this is typically not the case, current modeling developments in CASS include the extraction of environmental properties from measured reverberation. Future developments may reduce some of the known deficiencies.

10:05–10:20 Break

10:20

4aUW6. Sediment interface and volume reverberation modeling with the parabolic approximation. Kevin B. Smith, Lit-Siew Li, Boon-Chuan Lee, and Han Kao (Code PH/Sk, Dept. of Phys., Naval Postgrad. School, Monterey, CA 93943)

A new reverberation model based on the parabolic approximation is developed that includes sediment interface and volume perturbations. A multiple forward/single backscatter approximation is made, and the structure of the solution is found to depend on the two-way propagation with a scattering strength scaling dependent on the local properties of the perturbation. The model is implemented for continuous wave (CW) signals to predict mean reverberation pressure levels and for broadband pulse signals to generate complex reverberation structures in the time-domain. The spatial correlation and statistical properties of these predicted signals are then analyzed in an attempt to extract information on the underlying characteristics of the perturbation. Preliminary analysis suggests that reverberation due to the volume perturbations decorrelates more rapidly over depth than the reverberation due to interface fluctuations, although the differences appear small. Additionally, the statistical character of the reverberation structure due to the interface appears as a relatively flat spectrum, while the spectrum of the volume reverberation tends to appear colored. Attempts to correlate these characteristics with the structure of the perturbations is ongoing. [Work supported by ONR, Code 321OA.]

10:45

4aUW7. Reverberation and diffusion in layers with rough boundaries. David H. Berman (Dept. of Phys., Univ. of Northern Iowa, Cedar Falls, IA 50614, david.berman@uni.edu)

A pulse of sound is emitted from a point source located between two surfaces, one of which is rough. After a long time, sound that is returned to the source is governed by a diffusing remnant of the scattered sound. The diffusing remnant leads to algebraic, as opposed to exponential, decay of the reverberation when there is no absorption. An admittance formalism is used to describe this diffusive mode. Particular attention is given to the case in which the sound speed profile confines sound to the neigborhood of the

scattering surface. As in the case of light scattering by discrete particles, the frequency dependence of the scattering operator leads to a correction to the diffusion constant that is simply described in terms of the sound speed contrast between the region near the surface and far from the scattering surface. This result builds on a two-frequency generalization of the Ward identity for energy-conserving scattering.

Contributed Papers

11:10

4aUW8. Simulations of rough boundary effects on shallow water propagation. Eric I. Thorsos, Kevin L. Williams, Frank S. Henyey, Stephen A. Reynolds, and W. T. Elam (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

Results of PE simulations will be described for propagation in a shallow water waveguide with a rough sea surface and a flat seafloor. Pressure fields are simulated in two space dimensions and have been obtained using a wide-angle PE code developed by Rosenberg [A. D. Rosenberg, J. Acoust. Soc. Am. **105**, 144–153 (1999)]. At a frequency near 3 kHz and with typical sea conditions, the behavior of the forward scattered field is found to depend dramatically on the incident grazing angle. A narrow incident beam at a 10° grazing angle is completely broken up, with energy scattered widely in angle. At a 1° angle, little energy is scattered out of the beam. Examples will also be shown for the more typical case of a relatively broad beam with a wide range of grazing angles at the surface. At low frequency, such as 100 Hz, it is useful to represent shallow water fields in terms of normal modes. Results of PE simulations using single mode initial fields and sinusoidal surfaces will be used to illustrate the effects of surface scattering on mode propagation and on mode couplings. [Work supported by ONR.]

11:25

4aUW9. Numerical simulation of bottom reverberation in the time domain. Dajun Tang (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

A model that simulates a mid-frequency (1–4 kHz) direct-path reverberation field in shallow water is presented. This frequency range is important in the applications of detection sonar systems in shallow water. The model assumes that the scattering is caused by seafloor roughness and sub-bottom diffused heterogeneity. A Monte Carlo simulation of the reverberant field in the time domain based on the first-order perturbation approximation is developed. The roughness and sub-bottom heterogeneity are generated based on their power spectra. The computation is handled efficiently by taking advantage of the symmetry of backscattering geometry. The developed code can help to (1) investigate the relative importance of different scattering mechanisms, (2) predict spatial coherence of the scattered field, and (3) simulate and design field experiments. [Work supported by ONR.]

11:40

4aUW10. Multistatic active system performance modeling in shallow waters. David M. Fromm (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375-5320)

The Bistatic KRAKEN Reverberation (BiKR) model is an active system performance prediction model for bistatic source/receiver geometries based on adiabatic normal mode propagation [W. A. Kuperman, M. B. Porter, J. S. Perkins, and R. B. Evans, J. Acoust. Soc. Am. **89**, 125–133 (1991)]. By exploiting ray-mode analogies, it can provide detailed predictions of reverberation, target echo, and signal excess for low-frequency (<1 kHz) multistatic systems in range-dependent environments. This paper presents an overview of the models underlying principles and analysis capabilities such as the estimation of signal spreading loss. Data/model comparisons that required the inclusion of losses due to shear will be presented for a recent sea test on the Heceta Bank off the Oregon coast. [Research sponsored by ONR.]

Meeting of the Standards Committee Plenary Group

•**S1 ACOUSTICS**—U.S. Technical Advisory Group (TAG) for IEC/TC 29 Electroacoustics and ISO/TC43 Acoustics

•**S2 MECHANICAL VIBRATION AND SHOCK**—U.S. Technical Advisory Group (TAG) for ISO/TC 108 Mechanical Vibration and Shock as well as ISO/TC 108 SC1 Balancing, including balancing machines, SC2 Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures, SC3 Use and calibration of vibration and shock measuring instruments, SC5 Condition monitoring and diagnostics of machines and SC6 Vibration and shock generating system.

•**S3 BIOACOUSTICS**—U.S. Technical Advisory Group (TAG) for ISO/TC 43 Acoustics, IEC/TC 29 Electroacoustics, and ISO/TC 108/SC4 Human Exposure to Mechanical Vibration and Shock

•**S12 NOISE**—U.S. Technical Advisory Group (TAG) for ISO/TC 43/SC1 Noise

The meeting of the Standards Committee Plenary Group will precede the meetings of the Accredited Standards Committees S1, S3, and S12, which are scheduled to take place in the following sequence on the same day:

S12	9:45 a.m. to 11:45 a.m.
S1	2:00 p.m. to 3:15 p.m.
S3	3:45 p.m. to 5:00 p.m.

Discussion at the Standards Committee Plenary Group meeting will consist of national items relevant to all S Committees, plus a review of the international standardization (U.S. TAG) activities including reports on recent meetings and planning for forthcoming meetings.

Members of S2 Mechanical Vibration and Shock (and U.S. TAG for ISO/TC 108 and five of its Subcommittees, SC1, SC2, SC3, SC5, and SC6) are also encouraged to attend the Standards Committee Plenary Group meeting even though the S2 meeting will take place one day earlier, on Wednesday, 5 December 2001, at 8:30 a.m.

The U.S. Technical Advisory Group (TAG) Chairs for the various international Technical Committees and Subcommittees under ISO and IEC, which are parallel to S1, S2, S3, and S12 are as follows:

U.S. TAG Chair/Vice Chair	TC or SC	U.S. TAG
ISO		
P. D. Schomer, Chair H. E. von Gierke, Vice Chair	**ISO/TC 43** Acoustics	S1 and S3
P. D. Schomer, Chair H. E. von Gierke, Vice Chair	**ISO/TC 43/SC1** Noise	S12
D. J. Evans, Chair	**ISO/TC 108** Mechanical Vibration and Shock	S2
M. Schonfeld, Chair	**ISO/TC 108/SC1** Balancing, including Balancing Machines	S2
A. F. Kilcullen, Chair R. Eshleman, Vice Chair	**ISO/TC 108/SC2** Measurement and Evaluation of Mechanical Vibration and Shock as Applied to Machines, Vehicles and Structures	S2
D. J. Evans, Chair	**ISO/TC 108/SC3** Use and Calibration of Vibration and Shock Measuring Instruments	S2
D. D. Reynolds, Chair	**ISO/TC 108/SC4** Human Exposure to Mechanical Vibration and Shock	S3
R. L. Eshleman, Chair R. F. Taddeo, Vice Chair	**ISO/TC 108/SC5** Condition Monitoring and Diagnostic of Machines	S2
G. Booth, Chair	**ISO/TC 108/SC6** Vibration and Shock Generating Systems	S2
IEC		
V. Nedzelnitsky, U.S. TA	**IEC/TC 29** Electroacoustics	S1 and S3

Meeting of Accredited Standards Committee (ASC) S12 Noise

P. D. Schomer, Chair S12, and Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 43/SC 1, Noise
2117 Robert Drive, Champaign, Illinois 61821

R. D. Hellweg, Vice Chair, S12
Compaq Computer Corporation, Acoustics Lab, Mechanical Engineering Group, MR01-3/03, 200 Forest Street, Marlborough, Massachusetts 01752

H. E. von Gierke, Vice Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 43/SC 1, Noise
1325 Meadow Lane, Yellow Springs, Ohio 45387

Accredited Standards Committee S12 on Noise. Working group chairs will report on the status of noise standards currently under development. Consideration will be given to new standards that might be needed over the next few years. There will be a report on the interface of S12 activities with those of ISO/TC 43/SC1 Noise including plans for future meetings of ISO/TC 43/SC1. The Technical Advisory Group for ISO/TC 43/SC1 consists of members of S12 and other persons not necessarily members of the Committee. Open discussion of committee reports is encouraged.

Scope of S12: Standards, specifications and terminology in the field of acoustical noise pertaining to methods of measurement, evaluation and control, including biological safety, tolerance and comfort and physical acoustics as related to environmental and occupational noise.

Session 4pAAa

Architectural Acoustics and Speech Communication: Speech Intelligibility and the Metrics Used for its Evaluation II

Kenneth P. Roy, Cochair
Innovation Center, Armstrong World Industries, 2500 Columbia Avenue, Lancaster, Pennsylvania 17603

Peggy B. Nelson, Cochair
Communication Disorders, University of Minnesota, 115 Shevlin Hall, 164 Pillsbury Drive, SE, Minneapolis, Minnesota 55455

Chair's Introduction—1:00

Invited Papers

1:05

4pAAa1. General considerations regarding the acoustic comfort conditions in architectural spaces. Fausto Rodriguez-Manzo (Div. de CyAD, Universidad Autonoma Metropolitana, Azcapotzalco, Av. San Pablo 180, Edificio H, PB, Col. Reynosa Tamaulipas, 02200 Mexico, DF)

Recently the matter of acoustic conditions of various types of architectural spaces has been taking great importance. Such is the case of classrooms, open plan, and private offices. These types of rooms can be considered as noncritical acoustic rooms in opposition to theaters and concert halls that are traditionally considered as acoustic spaces. However, all kinds of architectural spaces can be considered as acoustic spaces. There are acoustically critical spaces and acoustically natural spaces. Acoustically critical spaces need specialized acoustic consulting. Acoustically natural spaces need criteria for acoustic comfort. Acoustic comfort is not a matter of precision, but a matter of criteria. This paper establishes the relation between architecture and the science of sound through the concept of acoustic comfort, enhancing the importance of the acoustic conditions of all types of architectural spaces. Acoustic comfort levels are theoretically examined according to several author conceptions, and finally an acoustic comfort criterion including speech intelligibility conditions for specific types of architectural spaces is presented.

1:30

4pAAa2. Merging EASE for sound reinforcement systems and CAESAR for room acoustics. Wolfgang Ahnert, Stefan Feistel, Bruce C. Olson (ADA Acoustic Design, Ahnert, Berlin, Germany), Oliver Schmitz, and Michael Vorländer (Institut für Technische Akustik, RWTH Aachen, Germany)

The combination of EASE with a state-of-the-art room acoustical simulation engine CAESAR, merging the advantages of both to form a multipurpose simulation tool, is introduced. The new tool is useful for consultants in room acoustics (ISO3382) and for the first time for electroacoustics as well. Advantages of combining a high quality speaker database and a powerful 3D-CAD interface including mapping presentations with a hybrid image source/ray tracing algorithm including diffuse scattering will be discussed. Details about the new EASE module CAESAR and its limitations as well as the extended possibilities of the system are the main topics of this presentation.

1:55

4pAAa3. Limitations of current sound system intelligibility performance measurement techniques and metrics. Peter Mapp (Peter Mapp Assoc., Colchester, UK, petermapp@btinternet.com)

The current and increasing trend to employ sound systems for emergency warning and evacuation purposes has led to an increased need to verify and assess the intelligibility performance of such systems. Applications range from passenger cabin PA systems in aircraft through to paging and emergency evacuation systems in industrial complexes, train stations, public buildings, and stadiums. Although a number of methods are available, these have either been borrowed from the fields of architectural and auditorium acoustics or one-dimensional communication networks. The paper shows that such approaches exhibit a number of fundamental inherent limitations that can lead to significantly reduced accuracy when assessing typical sound system performance. Although important factors such as background noise and reverberation can be accounted for, it is shown that the typical, nonlinear amplitude and frequency response characteristics of many systems are not adequately addressed. Furthermore, common temporal distortions such as discrete specular reflections or later arriving echoes are also shown to result in inaccurate assessments. The need for a simple, easy to use, and yet robust measurement technique is highlighted.

2:20

4pAAa4. Improving the intelligibility of aircraft PA systems. Peter Mapp (Peter Mapp Assoc., Colchester, Essex C03 4JZ, UK), Glenn Warnaka (Future Technologies, 1612 S. Allen St., State College, PA 16801), and Bruce Shimazu (C&D Aerosp., Huntington Beach, CA)

The speech intelligibility of commercial aircraft audio systems is of great importance. The authors have noted however, that subjectively, the intelligibility of such systems can vary considerably. Measurements made on a number of commercial flights indicate that signal to noise ratios are often around 0 dBA and frequently negative. As aircraft loudspeakers are located relatively close to the passenger, the collapsing narrower dispersion and beaming of the high frequencies from conventional cone loudspeakers can cause loss of intelligibility for passengers not on the axis of the loudspeaker and produce uneven sound distribution within the cabin. Distributed mode loudspeakers (DML) have demonstrated intelligibility superior to conventional loudspeakers in other applications such as public address systems and the opportunity was taken to evaluate their performance within an aircraft cabin environment. DMLs are very light in weight and can even be built into the internal trim, resulting in a near zero weight loudspeaker. Their essentially flat configuration also requires less volume. This paper presents a comparative evaluation of the speech intelligibility of current aircraft loudspeakers and DMLs as determined by both conventional word score tests and Speech Transmission Index measurements using both full STI and RaSTI. The tests were performed under both quiet and representative in-flight cabin noise conditions.

Contributed Paper

2:45

4pAAa5. Acoustic noise suppression for speech enhancement. David V. Anderson, Phil Spencer Whitehead, and Mark Clements (School of Elec. and Computer Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0250, spence@imtc.gatech.edu)

Removal of noise from a single channel audio signal is becoming an increasingly important problem due to the proliferation of portable communication devices. Early methods of single channel noise suppression introduced artifacts that could be as annoying as the original noise. Substantial effort has been focused on more effective noise suppression with fewer artifacts, but many difficulties remain and the problems seem far from being solved. For example, nearly all noise suppression techniques actually reduce the intelligibility of the processed signal. This paper investigates the sources of some of the artifacts common with Wiener filtering and proposes a structure to mitigate or eliminate these artifacts and effectively reduce noise without adversely affecting intelligibility. Results of one implementation of such a structure demonstrate significant improvements both in speech recognition performance and perceived speech quality.

Session 4pAAb

Architectural Acoustics: Measurements and Materials

David E. Marsh, Chair
Pelton Marsh Kinsella, 1420 West Mockingbird Lane, Suite 400, Dallas, Texas 75247

Contributed Papers

3:30

4pAAb1. A spherical microphone array-based system for the determination of spatial information in reverberant sound fields. Bradford N. Gover (Inst. for Microstructural Sci., Natl. Res. Council, Ottawa, ON K1A 0R6, Canada and Dept. of Phys., Univ. of Waterloo, Waterloo, ON N2L 3G1, Canada), James G. Ryan (Gennum Corp., Kanata, ON K2K 2A1, Canada), and Michael R. Stinson (Natl. Res. Council, Ottawa, ON K1A 0R6, Canada)

A measurement system has been developed for analysis of the directional and spatial variations in reverberant sound fields. Two spherical, 32-element arrays of microphones are used to generate narrow beams over two different frequency ranges, together covering 300–3300 Hz. Using an omnidirectional loudspeaker as excitation in a room, the pressure impulse response in each of 60 steering directions is measured. Through analysis of these responses, the variation of arriving energy with direction is studied. The system was first validated in simple sound fields in an anechoic chamber and in a reverberation chamber. The system characterizes these sound fields as expected, both quantitatively through numerical descriptors and qualitatively from plots of the arriving energy versus direction. The system was then used to measure the sound fields in several actual rooms. These sound fields were highly anisotropic, influenced greatly by the direct sound and early-arriving reflections. The rate of sound decay was not independent of direction, sound being absorbed more rapidly in some directions than in others.

3:45

4pAAb2. Personal computer and frequency analyzer measure reverberation time. Guillermo Arcas, Alberto Martn, and Cesar Asensio (Universida Politecnica de Madrid, Crta. Valencia Km-7, Madrid 28031, Spain)

This paper shows how a frequency analyzer (such as Brüel and Kjaer's Model 2144) can be used to perform reverberation time measurements with the help of some extra software. Although a particular solution is presented, the concept can be applied to any frequency analyzer that has a programmable interface. Brüel and Kjaer's Model 2144 is a powerful frequency analyzer, but it does not measure reverberation time. On the other hand, the analyzer has a programmable interface (GPIB) that permits us to control the instrument from a personal computer. This paper shows how a software application has been developed that gives the B&K2144 connected to a computer, the capacity of measuring reverberation time. The application has been developed using LabVIEW and it makes reverberation time measurements easy and reproducible.

4:00

4pAAb3. Estimated reverberation room absorption coefficients from an empirically derived transformation of predicted coefficients of large flat absorbers. Richard D. Godfrey (Owens Corning Sci. & Technol. Ctr., 2790 Columbus Rd., Rte. 16, Granville, OH 43023-1200, dick.godfrey@owenscorning.com)

ASTM C 423 specifies the industry accepted standard method to measure sound absorption coefficients by the reverberation room method. These coefficients are sometimes referred to as Sabin absorption coefficients. Because of the finite size of the sample used in these measurements, diffraction effects usually cause the apparent area of a specimen to be greater than the geometric area, thereby increasing the coefficients measured by this method over a sample of infinite lateral dimension which can be analyzed theoretically. In this paper, an empirically derived equation is present that transforms predicted absorption for large (infinite) flat absorbers to that which is measured by the reverberation room method. This empirically derived equation is based on an extensive database of measured absorption coefficients for A-mounted fiberglass materials with a wide range of flow resistivities and thickness. Estimated absorption is compared with measured data and infinite sample size predictions based on Mechel design charts [J. Acoust. Soc. Am. **83**, 1002 (1988)]. Comparisons are also made between the estimate made with this empirical method, and an analytical method, which accounts for diffraction developed by Northwood [J. Acoust. Soc. Am. **31**, 595 (1959); **35**, 1173 (1963)].

4:15

4pAAb4. Variation of attenuation constant and absorption coefficient of fiberglass with altitude. Mardi C. Hastings (Ohio State Univ., Dept. of Mech. Eng., 206 W. 18th Ave., Columbus, OH 43210, hastings.6@osu.edu)

ASTM C 423-99a specifies requirements for air temperature and relative humidity during decay rate measurements, but does not address variations in air density even though this parameter decreases with increasing altitude. The effect of changes in air density on absorption of fiberglass was investigated using approximations for characteristic impedance and complex propagation constant [M. E. Delany and E. N. Bazley, NPL AERO REPORT Ac37, March 1969], and design charts for absorbers in diffuse sound fields [F. P. Mechel, J. Acoust. Soc. Am. **83**, 1002–1013 (1988)]. At constant temperature and relative humidity, air density decreases by 17% from sea level to 5000 ft altitude, resulting in a 17% decrease in its acoustic impedance and a 12% increase in the attenuation constant of fiberglass. Using measured flow resistivities of four different fiberglass boards and an empirical correction factor to account for diffraction effects at sample edges, NRC values for the same board were found to vary within ± 0.04 between sea level and 5000 ft. Transmission measurements, such as those used to qualify aircraft insulation, will be overstated at higher altitudes because transmission loss is proportional to the attenuation constant. Thus correction factors should be developed for data taken substantially above sea level for comparison with data taken near sea level.

4:30

4pAAb5. Acoustic diffusers II. Alejandro Bidondo (Universidad de Buenos Aires, Eng. Faculty, A B Ingeniera de Sonido, Buenos Aires, Argentina)

Two lines of research are presented in this work about acoustic diffusers: A microscopic one which looks into its temporal operation presenting a new phenomenons vision, showing its memory property. The other, a pragmatic one, is very useful in the project advance control process, where a diffusion coefficient named Sound Field Diffusivity (SFD) is established for the quantitative evaluation of real diffuse fields. Further research suggestions are presented in both lines, one looking for a more complete surface diffusion coefficient that incorporates the time energy spreading, and on the other a frequency resolution of the SFD for precise design.

4:45

4pAAb6. A new technique of measurement of airborne sound insulation by means of adaptive processing. Antonio Minguez, Jorge Grundman, Juan Sancho, and Vladimir Ulin (EUIT Telecomunicacion, Ctra. Valencia km. 7, 28031 Madrid, Spain, aminguez@diac.upm.es)

In situ measurement of the airborne sound insulation, according to the ISO 140 Part IV specifications, needs the sound level pressures taken in the source room and in the receiving room, and the background noise level in this last room. Whenever the receiving room presents a great sound insulation, a high sound pressure level excitation is required in the source room in order to guarantee the appropriated margin over the background noise level. This higher sound pressure level is not ever attainable due to the limitations of the conventional sound sources. This procedure describes a new technique of measurement of the airborne sound insulation using a classical adaptive filter configuration that separates the background noise from the excitation sound in the receiving room so that a lower power excitation level can be used. Also, due to the possibility to obtain simultaneously and during a single measurement process the receiving sound pressure level and the background noise level, this technique allows to measure the actual background noise instead of a prior or post estimation of this magnitude once the receiving room level is measured.

5:00

4pAAb7. IMMOTUS*, sound insulation system. Daniel R. Schroer and Jean-Philippe Deblander (The Dow Chemical Co., Midland, MI 48674)

In 1999, a breakthrough sound insulation system was launched in the building industry. This system, developed with IMMOTUS*, a new polymeric foam from Dow, provides a unique answer to the critical issue of efficient sound insulation improvement with minimal space loss in dwellings and other building families. IMMOTUS R+ is a plasterboard laminate made of a 20 mm foam board and a gypsum board of 12.5 mm. When attached to a wall according to the system principle, a ceiling or a roof, this sound insulation system considerably reduces the airborne and impact noise transmissions. First in France and in the United Kingdom, then in other regions of the globe, a vast amount of data have been collected through Dow Laboratory experiments, Acoustic Institutes testing sessions, and successful commercial projects. During the course of this presentation, the originating concept and the main applications for this lightweight, high performance system will be described. Both the IMMOTUS* foam material and integrally associated systems properties will be reviewed. Acoustical behavior prediction and several configurations of full-scale experiments will be discussed.
(* Trademark of the Dow Chemical Company.)

THURSDAY AFTERNOON, 6 DECEMBER 2001 ROOM 315, 1:00 TO 1:30 P.M.

Session 4pABa

Animal Bioacoustics, Underwater Acoustics, Acoustical Oceanography and Noise: Acoustics and Marine Mammals III: Acoustics and Underwater Propagation II

Darlene R. Ketten, Chair
Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Contributed Papers

1:00

4pABa1. Morphometric analyses of hearing in two families of pinnipeds. Sarah E. Marsh, Iris Fischer, Julie Arruda, and D. R. Ketten (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Pinniped (seal and sea lion) auditory systems operate in two acoustically distinct environments, air and water. Otariids (sea lions and fur seals) generally divide their time evenly between land and water while phocids (true seals) spend the majority of their time in water. These pinniped families therefore offer an exceptional opportunity to investigate aquatic versus terrestrial hearing mechanisms. Recent reports indicate differences in hearing ranges and sensitivities among species from these families [D. Kastak and R. Schusterman, J. Acoust. Soc. Am. **1303**, 2216–2228 (1998); P. W. B. Moore and R. Schusterman, Marine Mammal Sci. **3**, 31–53 (1987)]. In this project, ear anatomy of three representative pinniped species (*Otariidae*: California sea lion, *Zalophus californianus; Phocidae*: northern elephant seal, *Mirounga angustirostris*; and harbor seal, *Phoca vitulina*) was examined using computerized tomography (CT scans) and gross dissection. Three-dimensional reconstructions of heads and ears from CT data were used to determine inter-aural dimensions and ossicular chain articulations. Ossicular weights and densities were measured using conventional and displacement techniques. Results strongly support a canal-centric system for pinniped sound reception and localization. Further, true seals show adaptations for aquatic high-frequency specialization. [Work supported by ONR.]

1:15

4pABa2. Temporary threshold shift (TTS) in bottlenose dolphins (*Tursiops truncatus*) exposed to tonal signals. James J. Finneran (Sci. Applications Intl. Corp., Maritime Services Div., 3990 Old Town Ave., Ste. 105A, San Diego, CA 92110), Donald A. Carder, and Sam H. Ridgway (Space and Naval Warfare Systems Ctr., San Diego, CA 92152-5000)

A behavioral response paradigm was used to measure underwater hearing thresholds in two bottlenose dolphins (*Tursiops truncatus*) before and after exposure to intense tones. The fatiguing sounds were 3 kHz, 1 s tones. Hearing thresholds were measured at 3 and 4.5 kHz using a vocal response technique and a modified version of the Method of Free Response. Tests were conducted in a 6 m diam, 1.5 m deep vinyl-walled swimming pool. Ambient noise spectral density levels in the pool were approximately 44 dB *re*: 1 Pa^2/Hz. Postexposure threshold measurements normally began within two minutes of exposure and continued for 1015 min following exposure. A temporary threshold shift (TTS) was defined as a 6 dB or larger increase in threshold compared to the corresponding pre-exposure level. Preliminary results indicate no TTS at 4.5 kHz for either subject at received SPLs of 200 dB *re*: 1 Pa. These data, as well as the test paradigm, absolute hearing thresholds, and observed behavioral reactions, will be presented and compared to existing TTS data for odontocetes.

Session 4pABb

Animal Bioacoustics, Underwater Acoustics, Acoustical Oceanography and Noise: Acoustics and Marine Mammals IV: Effects of Noise on Marine Mammals

Edmund R. Gerstein, Cochair
Leviathan Legacy, Inc., 1318 Southwest 14th Street, Boca Raton, Florida 33486

Joseph E. Blue, Cochair
3313 Northglen Drive, Orlando, Florida 32806

Chair's Introduction—1:30

Invited Papers

1:35

4pABb1. The Office of Naval Research program on the environmental consequences of underwater sound: Approach, results, and applications. Robert Gisiner (Office of Naval Res., 800 N. Quincy St., Arlington, VA 22217-5660)

The Office of Naval Research (ONR) is a basic and applied research funding organization established by the U.S. Navy in 1946 to stimulate research in a wide range of topics of relevance to naval and ocean science. In the mid- to late 1980s, ONR was one of the first research funding agencies to respond to the concern about possible adverse biological impacts from manmade underwater sound. Since 1991, ONR has had a large and growing research program on marine animal hearing and sound use, the effects of manmade sound on physiology and behavior, and the development of research tools for better assessing the effects of sound. Most recently, the program has increased emphasis on predictive models of underwater noise effects as a management and decision-making tool. The presentation will include specific examples of ONR-sponsored research and describe the mechanisms by which research results are made available to the government and the public.

1:55

4pABb2. The impacts of anthropogenic sounds on fishes. Arthur N. Popper (Dept. of Biol. and Neurosci. and Cognit. Sci. Prog., Univ. of Maryland, College Park, MD 20742, ap17@umail.umd.edu)

There is considerable concern that increases in marine sounds may be negatively impacting marine organisms. While most of the concern in the literature and in the press revolves around marine mammals, it is important to recall that the bulk of the marine vertebrate biota is made up of numerous fish species, and that many fishes are important in the food chains of marine mammals and humans. Moreover, since fishes use sound to learn about their environment and for communication, changes in ambient noise may have a significant impact on behavior and physiology of fishes, and thus on organisms higher in the food chain. Though there is a call to increase our knowledge of the impact of anthropogenic sounds on marine mammals, there have been few parallel studies looking at fishes (or on invertebrates and marine reptiles). This paper will review the little that is known about the impact of sound on fish behavior and physiology, and discuss the usefulness of these data in evaluating the impact of anthropogenic sounds on marine fishes. In addition, the paper will suggest a number of studies that are needed in order to have a better understanding of the impacts of anthropogenic sounds on fishes.

2:15

4pABb3. Parameters affecting seismic survey sound levels versus distance and marine mammal safety. Charles R. Greene, Jr., William C. Burgess (Greeneridge Sci., Inc., 4512 Via Huerto, Santa Barbara, CA 93110), and W. John Richardson (LGL Ltd., Environ. Res. Assoc., 22 Fisher St., King City, ON L7B 1A6, Canada)

Measurements of airgun sounds were made from 1996 through 1999 in the coastal waters north of Alaska (the Beaufort Sea). Sound levels versus distance (typically 100 m to *ca*. 6 km) from the source were measured for two single airguns and nine airgun arrays. Parameters of the measurements were number and size of airguns, array geometry and depth, receiver and water depths, and aspect. The total volume ranged from 10 to 120 $in.^3$ for the single airguns and from 56 to 1500 $in.^3$ for airgun arrays. The number of airguns ranged from 4 to 16. The U.S. National Marine Fisheries Service (NMFS) presently states that pinnipeds and cetaceans should not be exposed to received levels of pulsed sounds above 190 and 180 dB *re*: 1 μPa (rms over duration of pulse), respectively. NMFS often expresses concern that the disturbance may occur above 160 dB *re*: 1 μPa (rms). For each airgun source, distances within which animals might receive these three levels were determined. An effort is being made to relate the physical parameters of airgun operation to azimuthally radiated sound levels. [Work supported by BP Exploration (Alaska), Inc. and WesternGeco.]

2:35

4pABb4. Baleen whale responses to low-frequency human-made underwater sounds. Christopher W. Clark and Kurt M. Fristrup (Bioacoustics Res. Prog., Cornell Lab of Ornithology, 159 Sapsucker Woods Rd., Ithaca, NY 14850)

Baleen whales are vulnerable to impact from human-made underwater sounds. Most species produce communication calls and some sing, with most sound energy between 20 and 2000 Hz. Cochlear mechanics indicate inner ears well adapted for hearing below 1000 Hz. Many species breed and calve in coastal habitat, or feed along shelf edges or areas of ocean upwelling, and are frequently exposed to noise from commercial and recreational activities. Humans have become increasingly more adept at and dependent on exploring the ocean with acoustic probes. Sound sources are typically high intensity and in the primary acoustic production and perception frequencies of the baleen group. Evaluation of impact is difficult, confounded by a general lack of basic knowledge on baleen whale behavioral ecology, distribution, and abundance, and signal function. Results from three integrated research projects investigating whale responses to controlled exposures to Navy low-frequency sonar (LFA) indicate relatively low levels of short-term responses, even at received exposures as high as 150 dB re: 1 Pa. Results are interpreted relative to possible population level impact from an operational source. Results support a conceptual shift in impact from a single source, while emphasizing the need for a coherent, cautionary policy regarding cumulative and long-term impacts on the ocean environment. [Research supported by ONR.]

2:55–3:10 Break

Contributed Papers

3:10

4pABb5. Ocean ambient noise from anthropogenic and natural sources in the context of marine mammal acoustics. Douglas H. Cato (Defence Sci. and Technol. Organisation, P.O. Box 44, Pyrmont, NSW, Australia) and Robert D. McCauley (Ctr. for Marine Sci. and Technol., Curtin Univ. of Technol., Bentley, WA 6102, Australia)

The effect of anthropogenic ambient noise on great whales is difficult to assess because of the limited knowledge of the way in which these whales used sound. Some idea of the impact can be obtained by comparing anthropogenic noise with the natural ambient noise, i.e., the background noise that whales have always experienced. Most measurements of ambient noise, however, have been made in the northern hemisphere where there is significant contamination from anthropogenic noise. Around Australia, there are many areas where anthropogenic noise is low or negligible because of the relative isolation of the continent, and this has allowed us to make reliable measurements of the natural ambient noise throughout the frequency range of relevance to marine mammals, and in areas frequented by vocalizing whales. The results show that there is more than 30-dB temporal variation in natural ambient noise, with 20 dB being common. Much of the variation is due to the contribution of biological sources, especially fish and whale choruses, and sea surface noise (wind-dependent noise). Natural ambient noise levels are often comparable to those of traffic noise (noise from distant shipping), and at times reach the high-traffic noise levels of the northern hemisphere.

3:25

4pABb6. Modeling the noise exposure of marine animals during seismic surveys. Alec J. Duncan and Robert D. McCauley (Ctr. for Marine Sci. and Technol., Curtin Univ., GPO Box U1987, Perth, WA 6845, Australia)

This paper describes a technique that has been developed to estimate the noise exposure of marine animals during seismic surveys. A seismic source model was developed to account for the directionality of many-element sources and was used to compute the source output as a function of frequency and azimuth. Sea bed, water column, and bathymetry properties were established for inputting to an appropriate sound propagation model, in this case the Range Dependent Acoustic Model (RAM). RAM was run for representative propagation paths at a number of frequencies spaced at octave intervals. The results were approximated by simple functions to allow rapid calculation of the received sound level for a large number of shot location/receiver location combinations. For a large number of receiver locations the number of shots were determined for which sound levels exceeded species dependent thresholds. These results were grided on an appropriate grid covering the entire survey area and presented in several different ways to highlight both spatial and temporal variations.

3:40

4pABb7. Distributions and densities of marine mammals and their role in modeling acoustic effects. James H. Hain (Associated Scientists, Box 721, 3 Water St., Woods Hole, MA 02543)

Describing, modeling, and predicting acoustic effects on marine mammals depends on accurate inputs of various kinds. One required input is the distribution and density of marine mammals for a given area and time period. The quality of input is especially important for endangered species and critical habitats. Often, acousticians are requested to proceed with modeling in the absence of credible marine mammal data. This may produce unsatisfactory results for use in support of meeting environmental compliance and regulatory requirements. For any given period and area, it is unlikely that we will ever know the true population or density. Therefore, we are left to estimate it. Examples are provided of recent efforts to calculate marine mammal densities in two proposed Navy east coast shallow water training ranges. How many marine mammals of each species occupy a 500 square nautical mile area by season and depth strata? The effects of survey planning, sample size, data collection protocols, and analyses on the resulting estimates are shown. The inescapable conclusion is that increased interaction, collaboration, and understanding of methods and requirements between acousticians and survey biologists are not only desirable but essential.

Session 4pABc

Animal Bioacoustics, Underwater Acoustics, Acoustical Oceanography, and Noise: Acoustics and Marine Mammals V: Panel Discussion on ASA's Role in Marine Mammal Bioacoustics

Charles E. Schmid, Chair
Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, New York 11747-4502

A panel, consisting of representatives from a wide breadth of interests, will discuss ideas to improve standards, research, education and policy regarding current issues in marine mammal bioacoustics. Issues will include topics covered in joint sessions sponsored by Acoustical Oceanography, Animal Bioacoustics and Underwater Acoustics (3aUW, 3pAB, 4aABa, 4pABa, and 4pABb). The objective of this panel discussion is to begin dialogue on what may be the best role for ASA to serve as a resource on such complex issues as the effects of regulation on underwater acoustics experiments and the increase of ambient noise and active sonar signals on marine mammals.

Session 4pEA

Engineering Acoustics and Underwater Acoustics: Underwater Sensors and Sources

Leroy Dorman, Cochair
Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0220

Manuel T. Silvia, Cochair
SITTEL Corporation, 1206 Foothill Road, Ojai, California 93023

Contributed Papers

1:30

4pEA1. Dynamic structural-acoustic-piezoelectric finite element analysis of a layered 1-3 composite, half-wavelength resonator underwater transducer array. Robert M. Koch (Adv. Technol. Div. Naval Undersea Warfare Ctr., Code 8232, Bldg. 1302, 1176 Howell St., Newport, RI 02841, kochrm@npt.nuwc.navy.mil) and Kim C. Benjamin (Underwater Sound Reference Div. Naval Undersea Warfare Ctr., Newport, RI 02841)

The coupled structural, acoustic, and piezoelectric noise propagation physics within a layered 1-3 composite, half-wavelength resonator transducer array for high-frequency, broadband underwater applications is modeled herein using dynamic finite element analysis (FEA). This analysis is being performed primarily to evaluate the complex steady-state dynamic response of the multiple-layered composite array in order to ensure bond structural integrity during high-frequency operation. Specifically, various geometric configurations of the individual array layers will be examined to determine their effect on both array performance and structural integrity (e.g., relative PZT element spacing between layers, bond materials, joining layer thickness). The results of the present analysis will also be compared to experimental data being collected on prototype layered 1-3 composite arrays.

1:45

4pEA2. The design, fabrication, and acoustic calibration of a conical, octahedral array using 1-3 piezoelectric composite. Kim C. Benjamin and Steve E. Forsythe (Naval Sea Systems Command Div. Newport, Newport, RI)

Continuing to exploit the unique characteristics of 1-3 piezoelectric composite for wide-band acoustic array applications, a 120-element conical array was successfully fabricated and calibrated in less than 3 months. The eight-sided approximation of the conical aperture had dimensions 18.3 cm high, with a 13.3-cm base, and a 20 conical half-angle. Each of the 8 staves had 15 elements. The upper portion of the conical surface (cone tip) was not populated. The substrate thickness used in this demonstration was 1.8 cm providing a half-wave resonance frequency of approximately 86 kHz. The 1-3 piezocomposite configuration consisted of injection molded Navy type VI ceramic with a rigid epoxy back-fill that has proven to be both rugged and easily machined. The array was air backed. The array and tooling of the mechanical components were modeled using a computer-aided design package, and subsequently fashioned using stereo-lithography apparatus thus making this array a paperless design. Radiation modeling was accomplished using the Combined Helmholtz Intergral Equation Formulation (CHEIF). Measured results will include in-air laser vibrometer scans, and in-water transmit and receive beam patterns. [Work supported by Naval Sea Systems Command Div. Newport (Code 82).]

2:00

4pEA3. Oil-filled cymbal panels for acoustic projection applications. Thomas R. Howarth, Walter L. Carney (Naval Sea Systems Command Div. Newport, Newport, RI, howarthtr@npt.nuwc.navy.mil), James F. Tressler (Naval Res. Lab., Washington, DC), and Walter L. Carney (Naval Sea Systems Command Div. Crane, Crane, IN)

The U.S. Navy is developing cymbal-based elements arranged in a closely packed configuration for future underwater applications. This presentation concerns the design and experimental data of 49 elements placed within an oil-filled frame to form a 101 mm by 101 mm aperture. Experimental data shows a tunable low-frequency resonance between 5 and 10 kHz. The projector continues to increase in sound output to 100 kHz and then remains flat to 220 kHz. Higher overtones have been observed at 400 and 600 kHz. The units were tested under hydrostatic pressure from ambient to 800 psi and back. Laser Doppler vibrometry was performed over this frequency range to observe the modal behavior. The presentation will include a discussion of the results as well as planned future endeavors. [Work sponsored by ONR 321SS and ONR 321TS.]

2:15

4pEA4. The pressure gun, a wireline implosive source. Allan W. Sauter and LeRoy M. Dorman (Scripps Inst. of Oceanogr., Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0220)

We report development of an implosive source which is not based on stress failure. It is triggered by an electrical signal carried by armored cable, allowing precise timing. It can be used on the sea floor or in the water column. The trigger releases a small valve, allowing ambient pressure to operate a shuttle which opens a volume to flooding by seawater. It is recocked by returning to the surface. The strength of the source is best described by equivalent earthquake magnitude, which is proportional to volume change (a depth-independent quantity, in contrast with an explosive source). The 20-l device we have made has a moment magnitude of about -0.9. The duration of the signal is controlled by the time required to fill the volume, so the radiation of high frequencies is limited. This concentrates the energy radiated in seismically useful frequency band, and reduces effects on marine life. We have used this source to excite Scholte waves at a depth of just over 1100 m. At this depth the peak pressure occurs at about 20 Hz. It has the same source moment as a half-pound of explosive at about 700 m.

2:30

4pEA5. Extension of the long-line hydrophone calibrator to higher frequency. L. D. Luker (Naval Undersea Warfare Ctr. Div. Newport, USRD Code 216, Bldg. 1171B, Rm. 205, 1176 Howell St., Newport, RI 02841, lukerld@npt.nuwc.navy.mil)

The usable frequency range of the long-line hydrophone calibrator (LLHC) has been increased for endfire fields in a portion of the calibrator. The LLHC is a water-filled chamber for performing acoustic calibration of towed or other line arrays under operational environmental conditions. Projectors distributed along its length can produce an acoustic field equivalent to that seen by the array in open water from a sound source at any desired bearing and range. The original LLHC configuration was capable of measuring the sensitivity and directivity patterns of line arrays or modules up to 49 m long at temperatures from 2 °C to 35 °C, pressures up to 6.9 MPa, and frequencies from 5 Hz to 700 Hz. The LLHC has recently been modified in several ways to increase the usable frequency range to 2000 Hz for endfire fields over a 5 m long region inside the chamber. The modifications include changes to the way projector drives are computed, installation of thin-walled collars to make the acoustic impedance (and propagation speed) down the chamber more uniform, and the addition of four nonuniformly-spaced hydrophones. Results are presented that show maps of the sound field uniformity inside the calibration chamber. [Work supported by ONR Code 321.]

2:45

4pEA6. Theoretical and experimental investigation of acoustic dyadic sensors. Manuel T. Silvia (SITTEL Corp., 1206 Foothill Rd., Ojai, CA 93023) and Benjamin A. Cray (Naval Undersea Warfare Ctr., 1176 Howell St., Newport, RI 02841-1708)

The theoretical directivity index of a single combined acoustic receiver, a device that measures acoustic pressure, each component of particle velocity and velocity gradients at a collocated point, is presented. This receiver, described here as an acoustic dyadic sensor, is highly-directional, achieving as much as 9.5 dB of directivity gain. In practice, the higher-order velocity gradient components are obtained via a finite difference approximation of measured particle velocity. The statistical behavior of the errors in estimating the pure and mixed second-order partial derivatives are given. Experimental in-water measurements were completed on a prototype dyadic sensor, developed by EDO Corporation of Salt Lake City, Utah, and confirmed theoretical predictions.

3:00

4pEA7. Analysis and comparison of pressure gradient sensors of the fixed dipole and diffraction type with pressure gradient sensors of the motion accelerometer type. David A. Brown, Boris Aronov (BTECH Acoust., 1445 Wampanoag Rail, Ste. 115, E. Providence, RI 02915 and Elect. Eng. Dept., UMass Dartmouth, North Dartmouth, MA 02747), and Tetsuro Oishi (UMass Dartmouth, North Dartmouth, MA 02747)

Acoustic pressure sensors (hydrophones) and pressure gradient sensors are often used in combinations to characterize the acoustic sound field and to obtain compact devices with figure-of-eight and cardiod directivity patterns. Pressure gradient sensors of the fixed type can be realized by using pairs (doublets) of hydrophones spaced by a distance smaller than the acoustic wavelength or by using single elements whose size and geometry introduce a pressure difference from the diffraction of an acoustic wave. Pressure gradient sensors of the motion type are realized by using motion sensors such as accelerometers, velocimeters, or displacement sensors (seismometers) in a near neutrally buoyant package. Eulers' linear force equation confirms that the pressure gradient is proportional to the acoustic acceleration. We present our analysis of pressure gradient sensors of different types and show that a device of the fixed diffraction type can have the same sensitivity as a motion type accelerometer with comparable dimensions and using the same electroacoustic flexural disk transducer. Analysis of the response of sensors to unwanted actions is also presented. [Work supported in part by BTECH and ONR 321SS.]

3:15

4pEA8. Prototype high-sensitivity triaxial pressure gradient acoustic accelerometer. David A. Brown, Boris Aronov (BTECH Acoust., 1445 Wampanoag Rail, Ste. 115, East Providence, RI 02915 and Elect. Eng. Dept., UMass Dartmouth, North Dartmouth, MA 02747), Lawrence Reinhart, Glenn Volkema, and Tetsuro Oishi (UMass Dartmouth, North Dartmouth, MA 02747)

We present the results of the development of a high-sensitivity (400 mV/g) triaxial pressure gradient sensor of the acoustic motion type for underwater applications. The device uses piezoelectric flexural plates arranged symmetrically and mounted in a near neutrally buoyant case. The sensor is operated below resonance in the accelerometer mode. Hydrophones are attached to the case to realize a vector sensor that can produce a cardiod steerable in all azimuth and elevation. The four channels have a common acoustic center. Results of sensitivity and measured directivity beam patterns will be presented. [Work supported in part by BTECH and ONR 321SS.]

Session 4pMU

Musical Acoustics: General Topics in Musical Acoustics

Ian M. Lindevald, Chair
Science Department, Truman State University, Barnett Hall, Kirksville, Missouri 63501

Contributed Papers

1:30

4pMU1. Tone hole radiation directivity measurements. Gary P. Scavone (CCRMA, Dept. of Music, Stanford Univ., Stanford, CA 94305) and Matti Karjalainen (Lab. of Acoust. and Audio Signal Processing, Helsinki Univ. of Technology, FIN-02015 HUT, Finland)

Measurements have been conducted in an anechoic chamber for a comparison to current acoustic theory with regard to radiation directivity from a cylindrical pipe with toneholes. Several difficulties arise in measurements of this sort, including (1) the generation of sufficient driving signal strength at the pipe input for pickup by an external microphone; (2) external source-to-pickup isolation; (3) measurement contamination due to nonlinear driver distortion. Time-delay spectrometry using an exponentially swept sine signal was employed to determine impulse responses at points external to the experimental air column. This technique is effective in clearly isolating nonlinear artifacts from the desired linear system response along the time axis, thus allowing the use of a strong driving signal without fear of nonlinear distortion. The experimental air column was positioned through a wall conduit into the anechoic chamber such that the driver and pipe input were located outside the chamber while the open pipe end and tone holes were inside the chamber, effectively isolating the source from the pickup. Measured results are compared to both transmission-line, frequency–domain simulations, as well as time–domain digital waveguide calculations.

1:45

4pMU2. Time–domain synthesis of conical bore instruments. Gary P. Scavone (CCRMA, Dept. of Music, Stanford Univ., Stanford, CA 94305)

A series of approaches are presented for discrete time–domain synthesis of conical bore instrument sounds. This study uses a simple clarinet-like system, involving a memoryless nonlinear ''reed'' function and a distributed cylindrical air column model, as a point of departure for subsequent development. The generation of steady, self-sustained oscillations in such a system is complicated by properties inherent to truncated conical waveguides. Alternative methods include a structure equivalent to Benade's ''cylindrical saxophone,'' with two separate parametrization schemes. Finally, a more general ''virtual'' model is presented that is capable of synthesizing both half-wave and quarter-wave resonators. This structure provides a rich variety of possible sounds and offers an interesting perspective on conical waveguides.

2:00

4pMU3. Pitches, amplitudes, and centroids: The PACT model of sound segmentation. Scott Spiegelberg (Eastman School of Music, Rochester, NY 14607, spiegel@theory.esm.rochester.edu)

Logical sound segmentation is an essential tool in musical timber and articulation perception research. Unfortunately, the three current models of sound segmentation—the Constant Time Interval model, the 50% rms amplitude model, and the Amplitude/Centroid Trajectory model—have major theoretical flaws, preventing any real progress in these fields. The PACT (Pitch/Amplitude/Centroid Trajectory) model presents a new way to divide a musical sound into its attack, transition, steady state, decay, and release segments. This new method uses discrete wavelet transforms, spectral centroid envelopes and rms amplitude envelopes to define each of these segments. The most radical change is in the definition of the attack by its lack of pitch. The PACT model is supported by accepted theories of sound propagation and sound perception, with very clear results. The utility of the PACT model will be demonstrated with acoustical analyses of a variety of trumpet and violin articulations.

2:15

4pMU4. The influence of wavelets on music research. Scott Spiegelberg (Eastman School of Music, Rochester, NY 14507, spiegel@theory.esm.rochester.edu)

Starting in the late 1980s, wavelets have been used for a variety of purposes in music research. The earliest and still most prevalent use of wavelets has been time–frequency representations of musical signals, as a superior alternative to Fourier analysis. Computer music specialists interested in sound synthesis have also used wavelets as a means of dynamic filtering and separating monaurally superimposed sounds. One researcher has explored the use of wavelets in detecting patterns of whole pitches or rhythms, with varied results. And researchers at Yale University have used wavelets for noise reduction of a 1889 recording of Johannes Brahms performing his Hungarian Dance Number 1. In this paper the different types of wavelet transforms used in the music research mentioned above are explained, and future uses in timber research and sound synthesis is explored.

2:30

4pMU5. Wavelet-based automatic recognition of musical instruments. Bozena Kostek and Pawel Zwan (Sound and Vision Eng. Dept., Tech. Univ. of Gdansk, 80-952 Gdansk, Poland)

The objective of the present work is to automatically extract information from monophonic sounds. This process consists of several stages, namely, preprocessing, parameterization, and classification. This paper shows a thorough study on the wavelet-based parameterization of musical instrument sounds and automatic recognition by means of artificial neural networks (ANNs). First, an engineered method of pitch detection is presented and exemplified by several analyses. A short discussion on error associated with automatic pitch tracking is also included. Then, examples of time-frequency analyses of various musical instrument groups are presented. The analyses are performed employing a database containing musical sounds recorded at the Sound and Vision Engineering Department, Technical University of Gdansk. On the basis of such analyses a set of parameters is derived. Feature vector properties are then discussed. For that purpose Fisher statistics is used. It allows checking the separability between musical instrument pairs. In addition, for the purpose of automatic recognition of musical instrument groups artificial neural networks are used. Various structures and training methods of the ANNs are examined. Exemplary results obtained in the carried out investigations are provided and analyzed. Concluding remarks concerning further development of such experiments are also included in the paper.

2:45

4pMU6. Auditory roughness estimation of complex spectra—Roughness degrees and dissonance ratings of harmonic intervals revisited. Pantelis Vassilakis (Dept. of Ethnomusicology, Univ. of California, Los Angeles, 2539 Schoenberg Music Bldg., Box 951657, Los Angeles, CA 90095, pantelis1@msn.com)

Examination of musical instrument construction and performance practice indicates that sound variations involving the sensation of roughness are found in most musical traditions. A new roughness estimation model is proposed that better represents the theoretical knowledge and experimental results on sensory roughness. The model is based on previous studies [E. Terhardt, Acoustica **30**, 201–213 (1974); W. A. Sethares, *Tuning, Timbre, Spectrum, Scale* (Springer-Verlag, London, 1998)] and introduces an additional term to account for: (a) the contribution of the amplitudes of interferring sines to the roughness of a sine-pair and (b) the difference between amplitude modulation depth and degree of amplitude fluctuation. Compared to earlier roughness estimation models, the proposed model demonstrates the best agreement between estimated and observed roughness. Possibilities for further improvement are discussed. A hypothesis linking dissonance and roughness ratings of harmonic intervals within the chromatic scale is tested. Clear presence or absence of roughness appears to dominate dissonance ratings. In other cases, decisions on dissonance often ignore roughness and are culturally and historically mediated. The results suggest that, in the Western musical tradition, the consonance hierarchy of harmonic intervals corresponds to variations in roughness degrees. Further study should include cross-cultural empirical investigations.

3:00

4pMU7. Simultaneous performance of musical and verbal tasks. Barbara E. Acker-Mills (Dept. of Psych., Auburn Univ. at Montgomery, P.O. Box 244023, Montgomery, AL 36124-4023)

Allport, Antonis, and Reynolds [Q. J. Exp. Psych. **24**, 225–235 (1972)] demonstrated that pianists could sight read piano music and perform a verbal shadowing task. However, although widely cited in the multiple resource attention literature, the study has never been replicated. The current work provides a more detailed evaluation of the verbal component of the task and thorough statistical analyses. Effects of part and whole practice also are explored. Pianists completed a baseline phase where a Mozart minuet was performed and note/timing mistakes were recorded. A text passage from a SAT review book then was presented over headphones and reading comprehension questions were answered. In the dual task sessions, participants performed similar musical passages while listening to similar text passages. Reading comprehension questions were completed at the end of each dual task performance. Finally, half of the pianists practiced the musical and verbal tasks together for 10 min, while the other half practiced each individual task for 5 min each. Both groups repeated the dual task and the types of performance improvements were analyzed. Similarity to the Allport *et al.* results and effects of part and whole practice will be discussed.

3:15

4pMU8. Fractal dimensions of initial transients of musical instruments. Rolf Bader (Lindenallee 43, 20259 Hamburg, Germany, RolfBader@tiscalinet.de)

Fractal dimensions of initial transients are calculated for musical instruments. There are significant differences between a classical guitar, a violin, a clarinet, and a balinese Gender dasa. As the guitar and the clarinet have a maximum correlation-dimension of around $D=4$, the violin reaches up to $D=5.5$. The gender with its intensive inharmonic overtone structure can be up to $D=6.8$, and about $D=2.0$ within the steady state. Instruments with harmonic overtones always have $D=1$ during their steady state. These values seem to correspond to the number of rules governing the transient, which can be shown using zooming-in wavelet transform. In a sample of some hundred examined sounds, the guitar had just a light tendency for louder, higher, and Apoyando plucked tones to be more chaotic than for softer, lower-pitched, and Tirandu-played ones. The violin sounds differed extremely within these categories. The clarinet was also examined during pitch changes in melody. Here an information structure is proposed, which is the second derivation of time of a multifractal information-dimension and shows the stucture of information gain for these transients.

THURSDAY AFTERNOON, 6 DECEMBER 2001 ROOMS 118/119, 1:00 TO 2:00 P.M.

Session 4pNSa

Noise, Physical Acoustics and Animal Bioacoustics: Outdoor Sound Propagation in Coastal Areas

George A. Luz, Chair

U.S. Army Center for Health Promotion and Preventive Medicine, 5158 Blackhawk Road, Aberdeen Proving Ground, Maryland 21010-5403

Invited Papers

1:00

4pNSa1. Aircraft noise propagation in coastal areas. Alan F. Zusman (Dept. of Navy, Naval Facilities Eng. Command, Washington Navy Yard, Washington, DC 20374, zusmanaf@navfac.navy.mil)

Department of Defense (DoD) environmental noise management efforts employ computer models to project noise exposure around military installations, ranges, and operating areas. Both the DoD airfield noise model (NOISEMAP) and the heavy weapons noise model (BNOISE) assume propagation over land. In order to ensure the accuracy of these models for Navy and Marine Corps installations in coastal areas, the U.S. Navy has been funding studies of propagation over water. These studies have included: (1) joint work with the Swedish Air Force, (2) helicopter noise measurements along the southern California shore, (3) aircraft noise propagation studies in Florida and California, and (4) studies of gun fire propagation at the land–water boundary. Lessons learned from these efforts have been incorporated into the design of a new study of aircraft noise to be conducted this year.

1:20

4pNSa2. Guidelines for locating permanent noise monitors along a shore line. George A. Luz and Kimberly Estridge (U.S. Army Ctr. for Health Promotion and Preventive Medicine, 5158 Blackhawk Rd., Aberdeen Proving Ground, MD 21010-5403)

The performance of a noise monitoring system located along the shoreline of the upper Chesapeake Bay was analyzed by looking at measurements of a 3 lb. explosive charge detonated each work day morning at about 8:00 am. Monitors located (1) on a cliff (2) at the water's edge were more likely to register the explosions than monitors located as little as 500 m inland from the water's edge. These results were consistent with M.J. White's simulations of low frequency sound propagation across a land–water boundary [J. Acoust. Soc. Am. **108**, 2650 (2000)]. Comparison of measured levels with predictions from the Noise Assessment and Prediction System (NAPS) demonstrated a degradation of the NAPS predictions at the land–water boundary.

1:40

4pNSa3. Low frequency sound propagation across a land–water boundary. Michael White (Appl. Res. Assoc., Inc., Rocky Mountain Div., 5941 S. Middlefield Rd., Ste. 100, Littleton, CO 80123, mwhite@ara.com) and Larry Pater (Construction Eng. Res. Labs (CERL), P.O. Box 9005, Champaign, IL 61826-9005, L-PATER@cecer.army.mil)

Army noise assessment capability has recently been improved to account for the effects of land–water transitions on a received sound level. Sound waves propagating beyond a shoreline experience changes in surface impedance, surface roughness, and refraction profile. Measurements of the temperature and wind with tethersonde balloons indicate significantly weaker sound speed gradients over water than over land. When a temperature inversion was observed over land, the inversion was always smaller over the water. When the solar heating produced strong upward refraction near the land surface, the temperature profile over water was neutral (lapse). Simulations were run with a parabolic equation model incorporating range-dependent boundary refraction profiles and range-dependent boundary conditions. The signal levels exhibit changes with distance due to strong refraction over land, but the changes nearly cease in places where the propagation extends over water. Measured spectrum levels from explosion sound propagating across land–water boundaries agree well with the simulations. [Work supported by the USACERL, USACHPPM, and NAVFAC.]

THURSDAY AFTERNOON, 6 DECEMBER 2001 ROOMS 118/119, 2:15 TO 3:15 P.M.

Session 4pNSb

Noise: Noise in Dwellings, Communities and Annoyance

Angelo J. Campanella, Chair
Campanella Associates, 3201 Ridgewood Drive, Hilliard, Ohio 43026-2453

Contributed Papers

2:15

4pNSb1. What people want to hear. John J. LoVerde and Wayland Dong (Veneklasen Assoc., 1711 Sixteenth St., Santa Monica, CA 90404, jloverde@veneklasen-assoc.com)

As communities increase in population density, more multi-family dwellings appear in the marketplace. With the advent of acoustical legal action against developers and growing sophistication of owners, the need for the developer to communicate to the potential buyer the level of sound isolation provided has become important. The legal cases have created a renewed debate over the criteria and standards for these dwelling types. The Department of Housing and Urban Development (HUD) started the discussion of Acoustical Criteria in Multi-Family dwellings [U.S. Department of Housing and Urban Development, A Guide to Airborne, Impact, and Structure Borne Noise Control in Multi-Family Dwellings (1963)]. HUD later released minimum acceptable acoustical standards [U.S. Department of Housing and Urban Development, HUD Minimum Property Standards: Multi-Family Housing (1973)]. States, counties, and cities, through code requirements, have also provided minimum acceptable standards. This paper shall compare the existing standards. The comparison will include the types of noise sources and isolation systems that are addressed, and a discussion of the adequacy of the standards based on interaction with other consultants in legal cases.

2:30

4pNSb2. Mitigation of structureborne noise nuisance. Wing P. Ko (7312 S. Garnett Rd. E., Apartment No. 323, Broken Arrow, OK 74012)

Noise is radiated directly from the vibrating structure as structureborne noise. A structureborne noise case in a domestic building is described in this paper. The pipes were installed and mounted vertically with little or no insulation between the pipes and the walls. The result was that, every time a toilet was flushed or fresh water was used in one of the units in the building, the water flow and pump noise was heard clearly in the unit directly above the pump room. The family in the unit was very annoyed with the noise, especially in the middle of the night. The measured noise level inside the unit was found to exceed the statutory requirement of 45 dBA (Leq. 30 min) in the night time period (23:00–07:00 h). The solution to this problem was to isolate the pipes and pumps from the walls and floors with resilient materials or hangers to break the vibration channel.

Also, the pump operation was limited to only one time within any 30 min in the nighttime period. Finally, the noise level inside the unit was found to meet the statutory requirement.

2:45

4pNSb3. Survey and assessment of noise in coal preparation facilities. Jeffrey S. Vipperman (Dept. of Mech. Eng., Univ. of Pittsburgh, 531 Benedum Hall, Pittsburgh, PA 15228), Eric R. Bauer, Ellsworth R. Spencer, and Daniel R. Babich (Natl. Inst. for Occupational Safety and Health, Pittsburgh Res. Lab., P.O. Box 18070, 626 Cochrans Mill Rd., Pittsburgh, PA 15236)

Noise induced hearing loss (NIHL) has been found to be a significant problem in the mining industry. Between 70% and 90% of miners are reported to have enough NIHL to be classified as a disability (NIOSH, Publication No. 76-172, 1976; Franks, NIOSH Internal Report, 1996). In response to this problem, the National Institute of Occupational Safety and Health (NIOSH) is conducting a cross-sectional mining noise survey. Objectives for this survey include determining if various mining environments are within compliance with the recent mining noise regulation: 30CFR, Part 62 and if not, what engineering or administrative controls might be instituted. The preliminary results from the study of four coal preparation plants will be presented. Measurements included noise dosimetry in conjuction with time-at-task studies, 1/3-octave sound level measurements throughout each floor of the plants, and T_{20} reverberation time measurements. Results generally reflect fairly uniform A-weighted overall noise levels throughout the plant, which are capable of creating overexposures for a full 8-hour shift. The potential for administrative controls exists but the use of engineering controls is hindered by the size and number of the noise sources, the frequency content of the noise, accessibility requirements for the equipment, and industrial harshness of the environment.

3:00

4pNSb4. Annoyance of infrasounds. Ryszard Panuszka and Zbigniew Damijan (Staszic Univ. of Krakow 30 Al. Mickiewicza, Krakow PL 30-059, Poland)

At the present developed procedures of evaluation of annoyance of low frequency noise (LFN) on human and occupational health are based on international standards, ISO 7196 and ISO 9612. These standards are based on levels of thresholds of the human auditory system and subjective observations of vibrations of human body impacted by infrasound waves. In 1998 the first limits of exposure to infrasound and low frequency sounds were introduced to the guidelines of the American Conference of Governmental Industrial Hygienists (1999). Recent discoveries resulted significant sensitivity of the central nervous system on infrasonic waves especially below 10 Hz and the observed follow-up effects in the brain. New obtained results showed the impact of LFN on the electrical potentials in the brain accordingly to pressure of waves on the human body. Signals recorded in industrial workplaces were used for further studies with typical LFN spectra with 5 Hz in chamber at laboratory. External noise, electromagnetic fields, temperature, dust and other elements were controlled. Results show that follow-up effect is observed for 70% of the population. Multi-channel system of acquisition of signals was used for sound pressure levels, EEG-potentials, blood pressure monitoring, temperature, and pulse. Laboratory test included 45 volunteers of healthy people.

THURSDAY AFTERNOON, 6 DECEMBER 2001 ROOM 222, 2:15 TO 4:15 P.M.

Session 4pPA

Physical Acoustics: Outdoor Sound Propagation

Yves H. Berthelot, Chair

School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Contributed Papers

2:15

4pPA1. Nonlinear acoustic techniques for landmine detection: Experiments and theory. Murray S. Korman (Dept. of Phys., U.S. Naval Acad., Annapolis, MD 21402) and James M. Sabatier (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677, sabatier@olemiss.edu)

A nonlinear acoustic technique for detecting buried landmines has been suggested by Donskoy [SPIE Proc. **3392**, 211 (1998); **3710**, 239 (1999)]. Airborne sound at two primary frequencies $f1$ and $f2$ undergo acoustic-to-seismic coupling and a superimposed ''slow'' compressional wave interacts with the compliant mine and soil. The nonlinear mechanism involves a simple model of the top surface of the mine–soil planar surface separating two elastic surfaces. During the compression phase of the wave, the surfaces stay together and then separate under the tensile phase due to a relatively high compliance of the mine. This ''bouncing'' soil–mine interface is thought to be a bimodular oscillator that is inherently nonlinear. Geophone measurements scanning the soil's surface (at the difference frequency) profile the mine, but off the mine some nonlinearity exits. Amplitude-dependent frequency response curves for a harmonically driven mass-soil oscillator are used to find the nonlinearity of the soil acting as a ''soft'' spring. Finally, a layer of soil in a large tube, terminated by a thin elastic plate, is used to model the bimodular nonlinearity. [Work supported by the United States Army Communications-Electronics Command Night Vision and Electronic Sensors Directorate.]

2:30

4pPA2. Nonlinear seismo-acoustic land mine detection: Field test. Dimitri Donskoy, Alexander Ekimov, Nikolay Sedunov, and Mikhail Tsionskiy (Stevens Inst. of Technol., Hoboken, NJ 07030)

The paper presents results of the field test of the nonlinear seismo-acoustic technique for detection and discrimination of land mines [D. Donskoy, SPIE **3392** (1998); **3710** (1999)]. The tests were conducted in summer 2001 at the U.S. Army outdoor testing facilities. Plastic antitank mines (M19, VS1.6, VS2.2) and plastic antipersonnel mines (M14, VS50, TS50) were confidently detected at their maximum burial depths in both gravel and dirt lanes. Mine M14 is one of the smallest mine and practically undetectable by the competitive techniques, including NQR, GPR, etc. The test proved that the developed nonlinear seismo-acoustic detection approach is very sensitive to AT and AP mines, while completely insensitive to false targets, such as rocks, chunks of metal or wood, thus promising to deliver high probability of detection with low false alarm rate. The results of the tests are in good agreement with the developed physical model of the seismo-acoustic detection. [Work supported by the U.S. Army Research Office under Grant Nos. DAAG55-98-1-0402 and DAAD19-00-1-0039.]

2:45

4pPA3. Comparison of linear and nonlinear experiments for landmine detection. George Nock, Manaf Ali, James M. Sabatier (Natl. Ctr. for Physical Acoust., Coliseum Dr., University, MS 38677), and Murray S. Korman (Dept. of Phys., U.S. Naval Acad., Annapolis, MD 21402)

When airborne sound couples into the ground, so-called ''slow'' compressional waves can interact with a target buried in the soil and effect the vibration velocity of the surface. Donskoy [SPIE Proc. **3392**, 211 (1998); **3710**, 239 (1999)] has suggested a nonlinear technique that can detect an acoustically compliant buried mine that is insensitive to relatively noncompliant targets. A VS 1.6 inert antitank mine (23 cm diam) was buried in the natural loess soil at the test site. In the linear experiment, a pair of subwoofers were sinusoidally swept from 70 to 600 Hz. The geophone's response was measured along a 1 m scan segment across the mine. A 107 Hz resonance was chosen and the nonlinear experiment was performed by driving the woofers at $f1=300$ Hz and $f2=407$ Hz and detecting at the difference frequency. Nonlinear scan profiles over the mine have more relative sensitivity than the linear profiles. When a 30 cm diam, 4 cm thick concrete disk was buried, it could be detected ''linearly'' but not ''nonlinearly'' in agreement with Donskoy. Nonlinear theories will be discussed. [Work supported by the United States Army Communications-Electronics Command Night Vision and Electronic Sensors Directorate.]

3:00

4pPA4. Angle-of-arrival estimates for passive sensor arrays operating in atmospheric turbulence: Theory and experiment. Sandra L. Collier and D. Keith Wilson (U.S. Army Res. Lab., AMSRL-CI-EP, 2800 Powder Mill Rd., Adelphi, MD 20783-1197)

Angle-of-arrival (AOA) estimates are considered for a passive sensor array operating in atmospheric turbulence. The Cramer–Rao lower bounds (CRLBs) of the azimuth and elevation are calculated for a spherical wave assuming multiple unknown parameters, a line-of-sight propagation path, and a single monochromatic source. The received signal is modeled as a complex Gaussian random variable with a deterministic mean. Therefore, this theoretical treatment is valid for strong or weak scattering in the presence of strong diffraction. The experimental variances of the AOA estimates calculated using a spatial Fourier transform beamformer are compared to the theoretical CRLBs. The data were collected at a planar array under several distinct atmospheric conditions [D. K. Wilson *et al.*, Acoust. Res. Lett. Online **106**, L24–L29 (1999)]. The CRLBs are calculated for the meteorological conditions observed during the experiment. Models based on von Kármán's spectrum are used to model the atmospheric turbulence. The observed experimental standard deviations are typically found to be two to three times larger than the theoretical calculations.

3:15

4pPA5. Experimental tests of a generalized acoustic triangulation method. Jay E. Williams, Chad M. Williams, William G. Frazier (Miltec, Inc., NCPA, Coliseum Dr., University, MS 38677, jwilliams@mil-tec.com), and Kenneth E. Gilbert (Univ. of Mississippi, University, MS 38677)

A generalized acoustic triangulation method has been developed that takes into account the finite propagation speed of sound. The method has been tested using two 3-microphone arrays separated by 1 km, together with a low-flying commercial jet tracked via the Global Positioning System (GPS). The tests show that generalized triangulation is accurate as long as the bearing errors are small relative to the effective aperture of the two arrays. When the bearing errors are not small relative to the effective array aperture (e.g., at longer ranges), it is shown that significantly improved tracking accuracy can be obtained by assuming a constant speed target and using a bearings-only solution instead of triangulation. Wind noise is shown to be the major source of bearing errors. The degradation of bearing and tracking accuracy due to wind noise is demonstrated using field data, and possible approaches for wind-noise reduction are discussed. [Work supported by the U.S. Army Space and Missile Defense Command.]

3:30

4pPA6. The equivalent impedance of a comblike grating. Wenhao Zhu, Michael R. Stinson, and Gilles A. Daigle (Inst. for Microstructural Sci., Natl. Res. Council, Ottawa, ON K1A 0R6, Canada)

The scattering of acoustic plane waves by a comblike grating surface has been solved analytically. Both rigid and impedance gratings are considered. For the impedance grating, acoustic leakage through the partition walls is introduced by a leakage specific impedance. At low frequencies the grating can be represented by an equivalent impedance plane. Numerical results for the equivalent impedance are given for different grating geometries. It is shown that the equivalent impedance plane is locally reacting when the following conditions are met: $\lambda/a \gg 1$ and $l/a \gg 1$, where λ is the incident wavelength, a is the grating period, and l is the depth of the slots. The leakage impedance contributes to the nonlocal reaction of the equivalent impedance. When the incident wavelength is comparable to the grating period, a surface mode can be strongly excited, resulting in net surface wave propagation.

3:45

4pPA7. Modeling of acoustic pulse propagation over topographic ground. Lars R. Hole, Knut Lie, and Idar Dyrdal (Norwegian Defence Res. Establishment, P.O. Box 25, N-2027 Kjeller, Norway, lrh@ffi.no)

A hybrid coupled wave-number integration model and a time domain finite difference (TDFD) model have been used to predict propagation of low-frequency (below 500 Hz) acoustic pulses above ground with a hill present. The former model is also applied to study propagation along an upward slope. Ground conditions are described by poroelastic and viscoelastic theory. Air–ground interactions as well as the effect of important parameters such as permeability and thickness of poroelastic layers are discussed. Results are also compared with data from full-scale experiments.

4:00

4pPA8. Acoustic characteristics of porous media with tube-shape pores in water. Kang Il Lee, Heui-Seol Roh, and S. W. Yoon (Dept. of Phys., Sung Kyun Kwan Univ., Suwon 440-746, Republic of Korea, swyoon@skku.ac.kr)

Acoustic characteristics of nonrigid porous media with tube-shape pores in water are experimentally investigated. Lucite and sand sediment are selected as the specimens. The porosities of specimens are controlled with the artificial pores of circular cylindrical tube-shape. The fast and slow longitudinal waves are identified in such porous media. The phase velocity, pressure transmission coefficient, and attenuation coefficient of the fast and slow waves are measured as the functions of frequency and porosity. This study may be applicable to acoustic diagnosis of ocean sediment and cancellous bone. [Work supported by BK21 Program, KRF (KRF-2000-015-DP0718), and KOSEF (KOSEF-2000-0238-100) in Korea.]

Session 4pSA

Structural Acoustics and Vibration: Numerical Methods in Structural Acoustics II

John B. Fahnline, Chair
Applied Research Laboratory, Pennsylvania State University, 16 Applied Science Building, State College, Pennsylvania 16804

Invited Papers

2:00

4pSA1. Scattering from a coated cylindrical shell comparison of ray and exact computations. Andrew Norris (Dept. of Mech. and Aerosp. Eng., Rutgers Univ., P.O. Box 909, Piscataway, NJ 08854-8058)

We discuss sequentially the following: a homogenized shell theory and how it differs from uniform shell theory; high-frequency analysis of the quasi-P and S membrane wave solutions; the application of ray and coupling theory to the coated shell to predict the scattered response for oblique plane wave incidence; and numerical benchmarking against the exact code for thick coated shells. The shell theory is based on composite plate theory, with shell stresses modeled by the Kirchhoff assumption. The ray theory generalizes the theory developed for uniform thin shells, and accounts for the simultaneous P and S supersonic waves, and resulting resonances. The numerical comparisons demonstrate that the ray theory is accurate for undamped coatings, for which the excited resonances are strong. With damping present in the coating two unexpected discrepancies arise in the ray theory predictions. First, the damping of the excited structural waves is underestimated, even for relatively low damping in the coating, on the order of 0.1. At higher damping, e.g. 0.5, the same problem persists and is exacerbated by the degeneration of the background or specular approximation that is used in the asymptotic ray theory model. Refinements of the model to successfully address these is discussed.

2:30

4pSA2. A numerical technique for computation of flow-induced vibration and noise. Y. F. Hwang and Stephen Hambric (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804)

This paper discusses the use of commercially available structural acoustics solvers for computation the flow-induced vibration and noise at low speed. The major issue of this problem is the modeling turbulent boundary layer (TBL) forcing function for finite-element models of structures. Predictions of structural vibrations using the TBL cross-spectral density function directly is not computationally tractable for large complex structures, because extremely fine finite-element meshes had to be used to resolve the TBL pressure fluctuating scales. This paper discusses a more practical approximated discrete forcing function model, which represents only the low-wave number region of the TBL excitation in terms of the cross-spectral function of grid forces. The high-wave number (convective ridge) content of the excitation, which is insignificant for large structures with fixed boundaries, is thus ignored. This method is more feasible in practice because the finite-element meshes now need only resolve the structural wave scale instead of the convective turbulence scale. Numerical examples on the flow-induced vibration of a clamped plate, and the flow-induced noise radiated from a ribbed cylindrical shell will be discussed. [Work supported by ONR, Code 333.]

3:00

4pSA3. Prediction of elastic material properties via an adjoint formulation. Assad Oberai (Dept. of Aerosp. & Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215), Gonzalo Feijoo, and Peter Pinsky (Div. of Mech. and Computation, Stanford Univ.)

The prediction of elastic material properties via an adjoint formulation in this talk an efficient computational formulation to determine the acoustic properties of an elastic material, given its response to a time-harmonic excitation, is presented. Such problems arise in biomedical imaging using techniques such as Magnetic Resonance Elastography. The problem is posed as an inverse problem and its solution is computed using a mathematical programming algorithm (from mathematical programming). The functional to be minimized is the norm of the difference of the ''measured'' and a ''trial'' velocity field. At every iteration the trial velocity field is determined by solving the Helmholtz equation. The derivatives of the functional with respect to material properties, which form the input to the optimization algorithm, are computed by solving an adjoint problem. Using this formulation, the cost of computing derivatives is independent of the number of parameters used to represent the material properties, and thus the overall problem is rendered computationally tractable. A description of this methodology and an illustrative example to indicate its effectiveness will be presented.

3:30–3:45 Break

3:45

4pSA4. The "focusing" of transient response in structural networks: The development of "hot spots." John Ertel (Dept. of Phys., U.S. Naval Acad., Annapolis, MD 21402-5026) and Joe Dickey (Ctr. for Nondestruct. Eval., Johns Hopkins Univ., 3400 N. Charles St., MD 21218)

Vibration induced in a structural network will reverberate in and among the component systems and may, through constructive interference, create ''hot spots,'' points of anomalously high amplitude. Time domain responses of networks consisting of 2 and 4 one-dimensional wave bearing systems are calculated and displayed in the time/spatial domain and illustrate this phenomenon. Hot spots may be important in networks where they occur in fragile systems placing that system at risk for failure. Examples of physical structures where this model applies and numerical validation of this phenomenon are presented. An elementary illustrative example is a single physical beam supporting several wave types (thus several dynamic systems which interact with each other at the ends of the beam) driven, e.g., in compression but failing in flexure.

4:00

4pSA5. A new implementation of the finite elements method applied to vibro-acoustic analysis using object oriented programming. Pierre Lamary and Yvon Chevalier (ISMCM-CESTI, 3 rue Fernand HAINAUT, 93407 Saint-Ouen Cedex, France, pierre.lamary@ismcm-cesti.fr)

The aim of this paper is to present the most interesting points encountered in the creation of a new vibro-acoustic software. This program, named CAVOK (Computational Acoustics and Vibro-acoustics OK) applies specifically to the field of transportation (car, train, airplane). The originality of this work comes from the implementation of the finite elements method (FEM) proposed. Using an object oriented language (OOL) it has been shown possible to solve partial differential equations formally. First, it will be seen that the high level of abstraction proper to OOL facilitates software design. Secondly, we shall demonstrate the acuity of this approach for the treatment of acoustic insulating panels of aircraft. In such solutions, several physical domains interact, and must take into account the elastic behavior of the frame, the poro-elastic behavior of the absorbent material (glass wool), and the acoustic properties of air layers. The first results obtained are encouraging, and we shall conclude with the future research perspectives envisaged. In Particular, the CAVOK software will permit the comparison of different variational formulas of poro-elastic finite elements.

4:15

4pSA6. Acoustical crack detection of the structure: Wigner–Ville distribution versus short time Fourier transform. Leonid M. Gelman (Natl. Tech. Univ. of Ukraine, Kiev-103, P.O. Box 794/6, 01103, Ukraine, gelmanlm@mail.i.com.ua), Victor Giurgiutiu, and Adrian Cuc (Univ. of South Carolina, Columbia, SC 29208)

The innovative acoustical forced oscillation method with nonstationary excitation for fatigue crack diagnostics of the structure is considered. Because of the nonstationarity of the acoustical structure response, the modern time-frequency technique is used for diagnostics. The numerical simulation is carried out to investigate the diagnostic capabilities of the time-frequency signal processing technique based on Wigner–Ville distribution. Results from the Wigner–Ville procedure are compared to results from the short time Fourier transform. [Financial support of the U.S. National Research Council Twinning Program with Ukraine (Senior Program Officer Kelly Robbins) is thankfully acknowledged.]

THURSDAY AFTERNOON, 6 DECEMBER 2001 ROOMS 301/302, 1:30 TO 5:00 P.M.

Session 4pSC

Speech Communication: Production Models and Phonation (Normal and Disordered) (Poster Session)

Doug H. Whalen, Chair
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

Contributed Paper

All posters will be on display from 1:30 p.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:15 p.m. and contributors of even-numbered papers will be at their posters from 3:15 p.m. to 5:00 p.m.

4pSC1. Modeling tongue shapes with conic arcs. Khalil Iskarous, D. H. Whalen, and Ignatius Mattingly (Haskins Labs., 270 Crown St., New Haven, CT 06511)

Quantifying the changing shape of the tongue has been a longstanding challenge for experimental phonetics. One approach is to track the location of points on the tongue using x-ray microbeam or EMA. This leads to a precise numerical quantification, but is limited to information from a few points. Cinefluorography, MRI, and ultrasound remove the limitation by imaging entire sections of the tongue, but it is difficult to derive quantitative information from the scans. The present work illustrates a new method for quantifying an ultrasound tongue image by automatically fitting a conic arc to the midsagittal edge of the tongue. These arcs can vary continuously in shape from elliptic to parabolic to hyperbolic, and the variation is parametrized by two phonetically meaningful geometric parameters: one correlates with constriction location and the other with constriction degree. Quantification of tongue shape change is then measured by tracking the change in these 2 unique descriptors. One arc is sufficient for most tongue shapes, but some require two. Also, the general conic arc is a better model of the tongue dorsum than the circle, which has been used in articulatory synthesis systems. We will present data comparing these two models. [Work supported by NIH Grant DC-02717.]

4pSC2. A description of midsagittal tongue movement in terms of secant line kinematics. Richard S. McGowan (CReSS LLC, 1 Seaborn Pl., Lexington, MA 02420)

Tongue shape is a major determiner of the shape of the air channel in the vocal tract, and, hence, the acoustic signal during speech production. Flesh-point movement is a way that midsagittal tongue motion in speech production is examined, such as with the EMMA system and the x-ray microbeam system. We propose to transform the Cartesian coordinates of the flesh-points into one that specifies the position of the centers of the secant lines between each neighboring pair of flesh points, as well as the length and orientation of each secant line in this coordinate system, motion is considered in terms of translational, rotational, and stretching motions of the secant lines, with the former specifying gross motion and the latter two specifying local motion. Utterances from the X-Ray Microbeam Speech Production Database involving consonants with alveolar place of articulation are examined for kinematic differences resulting from differences in manner, vowel context, and syllable position. A further refinement of the secant line coordinates will be offered in terms of midsagittal secant line position and orientation relative to the palate and pharynx. This coordinate system should be helpful for interpreting flesh-point kinematics with regard to acoustic output.

4pSC3. FEM analyses of three-dimensional vocal tract models after tongue and mouth floor resection. Hironori Nishimoto, Masato Akagi (School of Information Sci., Japan Adv. Inst. of Sci. and Technol., 1-1 Asahidai, Tatsunokuchi, Nomigun, Ishikawa 923-1292, Japan), Tatsuya Kitamura (Faculty of Information, Shizuoka Univ., Shizuoka 432-8011, Japan), Noriko Suzuki, and Hiroto Saito (School of Dentistry, Showa Univ., Tokyo 145-8515, Japan)

The accuracy of vocal tract transfer functions estimated by Finite Element Method (FEM) analyses was evaluated to find relationships between structural characteristics of vocal tracts and their acoustical characteristics. Three-dimensional vocal tract models were constructed from MR Images of four speakers (two normal persons and two patients after tongue and mouth floor resection), uttering the Japanese vowel /i/. Vocal tract transfer functions of the models were estimated by FEM and one-dimensional models, and peak positions of these functions were compared with formant frequencies of these speaker's speech waves. The patient speech waves had a peculiar formant at around 1500 Hz. The peak positions of the patients estimated by FEM were similar to the formant frequencies of the patients, especially the formant at 1500 Hz. However, the transfer functions by the one-dimensional models had no peak at around 1500 Hz and the number of peaks was fewer than that by FEM. Since the patient vocal tract shapes were not symmetrical and bent right and left, sound pressure distributions estimated by FEM showed that the waves in the patient vocal tracts cannot be assumed as plane waves. Then, the one-dimensional models cannot describe acoustical characteristics of the patient vocal tracts.

4pSC4. Intrinsic *F*0 and tongue depth in ATR languages. D. H. Whalen (Haskins Labs., 270 Crown St., New Haven, CT 06511) and Bryan Gick (Univ. of British Columbia, Vancouver, BC V6T 1Z1, Canada)

Intrinsic *F*0 (IF0) is the tendency for high vowels to have a higher *F*0 than low vowels. While IF0 has been found for every language studied [D. H. Whalen and A. Levitt, J. Phonetics **23**, 349–366 (1995)], none of those languages distinctively contrasted advanced tongue root (ATR). Most languages conflate high tongue position with advanced tongue root, but some distinguish vowels by ATR. Here we report on two such languages, Igbo (a Niger-Congo language of Nigeria) and Kinande (a Bantu language of Congo). Acoustic analysis of *F*0 in Igbo suggests that +ATR vowels have higher *F*0s than their ATR counterparts. For Kinande, a combined acoustic and ultrasound study was performed. All five vowels /ieaou/ occur in ±ATR versions. All +ATR vowels except /i/ had higher *F*0s than their ATR counterparts. Horizontal depth at tongue root (distance from the jaw to the anterior pharyngeal wall) varied with ATR, as expected, and correlated with *F*0. Height also affected *F*0, in an amount of the same magnitude as for ATR. Some of IF0 appears to be attributable to pharyngeal shape (inferred from tongue depth), not to tongue height. Further work will examine which plays the more important role. [Work supported by NIH Grant DC-02717.]

4pSC5. Rear pharyngeal wall placement during vowel production. Harriet S. Magen (Haskins Labs., 270 Crown St., New Haven, CT 06511 and Dept. of Commun., Rhode Island College, Providence, RI 02908) and A. Min Kang (Haskins Labs., New Haven, CT 06511)

While the rear pharyngeal wall has been assumed to be stationary during speech, this needs to be confirmed. The present study used midsagittal magnetic resonance images to determine whether the posterior pharyngeal wall shows vowel-specific movement during the sustained production of 11 English vowels. Distances were measured between the most posterior portion of the vertebral body and the pharyngeal wall at four points. The first traversed the junction between the dens and body of the second vertebra (C2). The next three traversed the inferior border of the bodies of C2, C3, and C4. These ranged from a level just below the velopharyngeal port to one near the larynx. In preliminary measurements for one speaker, only the uppermost location showed significant variation due to vowel quality; at the other locations, differences were less than 1 mm. Further subjects are being studied. These findings confirm, at least for vowels, the assumption that the rear pharyngeal wall is stationary during speech. Therefore, midsagittal distances in the pharyngeal region can be inferred from measurements of the anterior pharyngeal wall in relation to the fixed features of the vocal tract. [Work supported by NIH Grant DC-02717.]

4pSC6. Intrinsic pitch in German: A puzzle? Christine Mooshammer[a] (Res. Ctr. of General Linguist., Jaegerstr. 10/11, 10117 Berlin, Germany, timo@zas.gwz-berlin.de), Phil Hoole (Inst. fuer Phonetik, Muenchen, Germany), Peter Alfonso (Office of Res., Knoxville, TN 37996-0140), and Susanne Fuchs[b] (Res. Ctr. of General Linguist., Germany)

Intrinsic pitch (IF0) differences have often been attributed to a physiological consequence of articulation, i.e., higher tongue-jaw position yields higher *F*0. However, Fischer–Jorgensen showed that German tense-lax pairs are produced with a similar IF0 despite the lower tongue positions of lax vowels. She also found better agreement between IF0 and jaw position than for tongue height. The first aim of this study was to replicate Fischer–Jorgensen's study using EMMA (seven speakers). Second, the contribution of jaw position on IF0 was examined by a bite-block condition (three speakers). The results of Fischer–Jorgensen were confirmed concerning similar IF0 of tense-lax pairs, where the jaw rather than the tongue better correlated with IF0, because jaw position differences between tense-lax vowel pairs were smaller than the tongue position differences. However, it seemed unlikely that this could be any kind of casual explanation, because frequent cases were found of somewhat higher *F*0 in the lax member of a pair, despite somewhat lower jaw position, and because of the fact that the same patterns of IF0 differences were found for the bite-block and normal condition. [a] Currently at Dept. of Audiol. and Speech, University of Tennessee, Knoxville, TN. [b] Also at Queen Margaret Univ. College, Edinburgh.

4pSC7. A distinctive region model based on empirical vocal tract area functions. Brad H. Story (Univ. of Arizona, Speech and Hearing Sci., P.O. Box 210071, Tucson, AZ 85721-0071)

The development of the Distinctive Region Model (DRM) [Mrayati *et al.*, Speech Commun. (1988)] is based on theoretically derived acoustic characteristics for a tube of uniform cross-sectional area assumed to approximate a neutral vocal tract configuration. Formant sensitivity functions calculated for the uniform tube are used to divide the vocal tract into

distinctive regions that, when constricted or expanded, will cause the formant frequencies to change in a predictable pattern. This study compares the original DRM (based on a uniform tube) with a new version created from a neutral vocal tract area function derived from published MRI data. Because it is subject to physiologic constraints, this neutral area function is nonuniform in cross-sectional area variation but exhibits formant frequencies similar to a uniform tube. Sensitivity function calculations for *F*1, *F*2, and *F*3 also show similarities to those of a uniform tube, but the zero-crossing points that divide the vocal tract into distinctive regions are shifted toward the lips. The result is distinctive regions that are not symmetric about the vocal tract mid-point but rather the back and front regions occupy about 60% and 40% of the total tract length, respectively. [Work supported by NIH R01-DC04789.]

4pSC8. Effects of false vocal fold width on translaryngeal flow resistance. Meena Agarwal and Ronald C. Scherer (Dept. of Community Disord., Bowling Green State Univ., Bowling Green, OH 43403)

Our purpose in this study was to determine the effects of the false vocal fold width (FVFW) on laryngeal flow resistance. False vocal folds were placed within a Plexiglas model of the larynx. The dimensions and placement of the false vocal folds were based on an earlier anatomical study of laminagrams during phonation. The ratio of the FVFW and the minimum glottal diameter (Dg) used in this study ranged from 0.60 to 21.5. Results suggest that for large FVFW/Dg values, flow resistance is not affected. For ratios near 1 and lower, the resistance is greater than for the true folds alone. For intermediate ratio values, the flow resistance is less than for the true folds alone. These intermediate values should be prevalent during maximal glottal opening. The results obtained so far are dependent upon glottal diameter, with a greater reduction in flow resistance for larger values of glottal diameter. The results suggest that the false vocal fold width may help control laryngeal airflow differently throughout the phonatory cycle, and should be considered in models of phonation.

4pSC9. Flow separation in a forced oscillation glottal model. Fariborz Alipour (Dept. of Speech Pathol. and Audiol., Univ. of Iowa, 334 WJSHC, Iowa City, IA 52242, alipour@shc.uiowa.edu) and Ronald C. Scherer (Dept. of Commun. Disorder, Bowling Green State Univ., Bowling Green, OH 43403, ronalds@bgnet.bgsu.edu)

Flow separation was investigated in pulsatile flow within a computational fluid dynamics (CFD) model of glottis motion. Two-dimensional time-dependent laminar flow was calculated for given dynamic wall geometries and flow boundary conditions. A Lagrangian–Eulerian method with a multiblock mesh was employed to allow grid points to move in space to conform to the boundaries. The glottal wall was approximated with two sinusoidal curves (one for the glottal entrance and one for glottal exit) connected with a straight line. The entrance and exit glottal diameters were changed as a sinusoidal function of time with a phase difference to generate the desired glottal wall movements. Results will be reported on the mean velocity and pressure distributions for various phase angles, frequencies, oscillation amplitudes, and Reynolds numbers. The locations of the flow separation points were obtained through a shear stress analysis using MATLAB. An increase in the glottal amplitude causes the separation points to move upstream. The study provides a better understanding of the effect of inferior–superior phase differences on the glottal flow. [Work supported by NIDCD Grant No. 1 R01 DC03566.]

4pSC10. Empirical measurements of vocal fold viscoelasticity at relatively high frequencies. Roger W. Chan (Audiol. and Speech Sci., Purdue Univ., West Lafayette, IN 47907)

Viscoelastic shear properties of human vocal fold tissues have been quantified in details in a series of previous experiments [see, e.g., Chan and Titze, J. Acoust. Soc. Am. **106**, 2008–2021 (1999)]. However, due to mechanical limitations of the rheometer used for empirical viscoelastic measurements, data have only been obtained at a relatively low frequency (up to 15 Hz). These low-frequency data are an order of magnitude below the physiological frequency range of phonation (on the order of 100–1000 Hz), requiring data extrapolation based on various theoretical approaches. The only truly valid approach, however, is direct empirical measurement of viscoelastic data at high enough frequency. Using a controlled strain rotational rheometer free of the inertial problems present in controlled stress rheometers, this paper reports preliminary measurements of dynamic shear data made at relatively high frequencies (up to 80 Hz), close to the frequency range of phonation in male speech.

4pSC11. Analysis of factors affecting vocal fold impact stress using a mechanical model. Heather E. Gunter, Robert D. Howe (Div. of Eng. & Appl. Sci., Harvard Univ., 29 Oxford St., Cambridge, MA 02138, gunter@fas.harvard.edu), Robert E. Hillman (Mass. Eye & Ear Infirmary, Boston, MA), and Kenneth N. Stevens (MIT, Cambridge, MA 02139)

High mechanical stress levels in the vocal folds likely contribute to pathological changes to the tissue such as vocal nodules. Vocal fold collision during phonation was modeled as a dynamic contact problem in order to calculate stress levels during impact. The model incorporated a 3-D, linear elastic, finite element representation of a single vocal fold, a rigid midline surface, and a simplified aerodynamic waveform. The effects of fold geometry, material properties, and sub-glottal pressure on peak impact stress were calculated. Impact stress time course and the relationship between sub-glottal pressure and peak impact stress agree with published experimental measurements [Jiang and Titze, J. Voice **8**, 132–144 (1994)]. Identification of high-stress configurations will promote better understanding of the etiology, persistence, and treatment of certain pathological voice changes. [Work supported by the Whitaker Foundation.]

4pSC12. Intraglottal pressures for large glottal diameters. Ronald C. Scherer, Kenneth J. De Witt (Dept. of Commun. Disord., Bowling Green State Univ., Bowling Green, OH 43403), and Bogdan R. Kucinschi (Dept. of Mech., Industrial, and Manufacturing Eng., Univ. of Toledo, 4006D Nitschke Hall, Toledo, OH 43606-3390)

This study examined the glottal wall pressures for larger glottal diameters, for which the glottis can not be considered ''long.'' A Plexiglas model was used. The glottal duct length was held constant at 0.30 cm, and the ratios of glottal diameter to duct length were 0.267, 0.533, and 1.07 (diameters of 0.08, 0.16 and 0.32 cm, respectively). Fourteen pressure taps lined the vocal fold surfaces. Diameters ranged from divergent 40° to convergent 40°. Transglottal pressures ranged from 1 to 15 cm H_2O. Results suggest that (1) for diffuser shapes, greater diameters tended to move the minimum glottal pressure upstream of the minimum diameter, and could enhance the pressure differences between the two sides of the glottis; (2) for the uniform glottis, for a diameter-length ratio of 0.533, the normal pressure decrease along the glottis did not occur, and for the ratio of 1.07, created a pressure rise (not fall) along the glottis; (3) for the 1.07 ratio for the converging glottal shapes, there were bimodal pressure profiles with pressure dips near the glottal entrance and the glottal exit. These findings are important in model building and determining the air pressure forces during phonation near maximum glottal excursion. [Work supported by NIH.]

4pSC13. Three-dimensional reconstruction of vocal fold geometrical deformation induced by formalin fixation. Miwako Kimura, Niro Tayama (Dept. of Otolaryngol., Univ. of Tokyo, Tokyo, Japan), Roger W. Chan (Purdue Univ., West Lafayette, IN 47907), and Ingo R. Titze (Univ. of Iowa, Iowa City, IA 52242)

In our previous study [Kimura *et al.*, J. Acoust. Soc. Am. **109**, 2413 (2001)], we showed that significant tissue deformation and distortion of vocal fold geometry is induced by the commonly used histological proce-

dure of formalin fixation. Significant shrinkages of the vocal fold body (thyroarytenoid muscle) and considerable distortion of the vocal fold cover (particularly the free edge) were observed. In mid-membranous coronal sections of canine tissues, there was an average of 20% shrinkage in vocal fold thickness and 40% shrinkage in vocal fold depth. This paper reports a three-dimensional (3-D) follow-up to these two-dimensional measurements. Vocal fold coronal sections every 2 mm along the anteroposterior direction were prepared from the same canine tissue samples, allowing for 3-D reconstruction of geometrical deformation. Results showed that the free edge of the vocal fold cover was significantly deformed, as emphasized by 3-D images.

4pSC14. Fusion of aspiration noise in voice synthesis. Blas Payri and Jody Kreiman (UCLA, Bureau of Glottal Affairs, 31-24 Rehab, Los Angeles, CA 90095-1794, blas@limsi.fr)

Synthesizing natural-sounding aspiration noise and breathiness remains challenging, because added noise may be masked or may fail to fuse with the harmonic component of the voice source, resulting in the perception of two separate sound channels. This study investigated the conditions leading to fusion and masking of aspiration noise in voice synthesis. Sustained vowels were synthesized using combinations of three noise-to-signal ratios, three $F0$ levels, two different versions of the harmonic component of the source, three sets of formants, and three differently shaped noise spectra. For each sample, expert and naive listeners judged whether they heard breathiness, and whether this breathiness noise was part of the voice or separate from it. Effects of the different experimental variables on masking and fusion will be discussed in an auditory scene analysis framework, combined with the perception of speaker naturalness.

4pSC15. Phonation types in Suai. Arthur S. Abramson (Haskins Labs., 270 Crown St., New Haven, CT 06511-6695, abramson@uconnvm.uconn.edu) and Theraphan Luangthongkum (Chulalongkorn Univ., Bangkok 10330, Thailand)

Many Mon-Khmer languages have such phonologically distinctive phonation types as modal, breathy, and creaky voice. The Suai language studied here is spoken in northeastern Thailand. Literature exists on these phonation types (''voice registers''), including some attention to acoustics; however, no perceptual studies appear to have been published. Suai was chosen here because it has just two contrasting categories conventionally labeled modal voice versus breathy voice. The Klatt synthesizer was used to control five of its parameters: $F0$, amplitude of turbulence, open quotient of the voicing waveform, overall gain, and spectral tilt. All combinations of these parameter settings yielded 51 variants of a basic syllable. These stimuli were identified by native speakers as one word or the other of a minimal pair. The results showed breathiness and pitch to be major cues. The original recordings of our informants, however, appeared only to have a pitch difference, an impression tested by acoustic analysis. A reliable difference was found only in $F0$ contours, with a lower one on the ''breathy'' type. There appears to be evidence for an ongoing shift from a system of phonation types to one of tones. [Work supported by NIH Grant DC-02727 and the Thailand Research Fund.]

4pSC16. Three-way contrast in breathiness in Ju—'hoansi. Amanda Miller-Ockhuizen (Dept. of Linguist., Cornell Univ., 203D Morrill Hall, Ithaca, NY 14853-4701)

I present results of an acoustic study investigating three female and three male speakers' productions of a three-way contrast in breathiness in the Khoisan language Ju—'hoansi, with the spectral slope (H1–H2) occurring at the release of the stop resulting in aspirated consonants, or the peak H1–H2 reached over the first part of the vowel in partially breathy vowels, or the peak H1–H2 maintained over the entire vowel in fully breathy vowels. H1–H2, and gamnitude of the first rahmonic peak were measured throughout the course of the vowel using a window of 46.5 ms, with a sampling rate of 22 050 Hz, starting from the first computer-recognized glottal pulse, with a step-size of 10 ms. Results show extremely similar patterns for both H1–H2 and gamnitude of the first rahmonic peak for aspirated consonants and partially breathy vowels. These sounds are distinguished on the basis of VOT, but not by fundamental frequency ($F0$). Aspirated consonants and partially breathy vowels are extremely similar in H1–H2, gamnitude of the first rahmonic peak, and $F0$, being differentiated only by voice onset time (VOT).

4pSC17. Perception of tremor rates in pathological voices. Jody Kreiman and Bruce R. Gerratt (Division of Head and Neck Surgery, UCLA School of Medicine, 31-24 Rehab Ctr., Los Angeles, CA 90095-1794)

Vocal tremors characterize many pathological voices, but acoustic-perceptual aspects of tremor are poorly understood. To investigate the perceptual importance of the rate of frequency modulation, we created synthetic copies of ten pathological voices. For half the voices, tremors were approximately sinusoidal, and for the other half, tremors were irregular in rate and amplitude. Tremor rates in the copies were varied in nine steps of 0.25 Hz. Listeners heard each synthetic copy paired with the natural voice sample, and judged which of the copies best matched the target voice. Small differences in tremor rates were easier to hear when tremors were sinusoidal than when they were irregular, but overall, listeners were not especially confident of their judgments. For sine wave tremors, differences in rate were easier to hear for faster rates, but wider extents of tremor significantly reduced the ability to detect rate differences. Similar effects of rate manipulations were not observed for irregular tremors. These results suggest that the perceptual importance of tremor rates depends on the total pattern of frequency variation. In the most complex cases, listeners apparently respond to overall levels of $F0$ variability rather than to precise patterns of frequency change. [This research was supported by the NIH.]

4pSC18. Prosodic differences in tracheoesophageal speech: Perception and production. Kimberly Rosenbauer and Michael Trudeau (Dept. of Speech and Hearing Sci., The Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210-1002, rosenbauer.2@osu.edu)

Prosody plays an important but often unrecognized role in a listeners interpretation of an utterance. The potential for significant prosodic limitations is often found in speakers who use tracheoesophageal speech due to extirpation of the larynx from cancer. This is because many prosodic features in the normal speaker are controlled by the use of laryngeal muscles [Kent, *The Speech Sciences* (1997)]. Previous studies showed inconsistencies in intonation and word stress perception and production levels regarding tracheoesophageal speech [Gandour and Weinberg, J. Speech Hear. Res. **26**, 142–148; Trudeau *et al.*, J. Commun. Dis. **21**, 21–31]. The present study was designed to reevaluate listeners perceptions of tracheoesophageal speech and then describe the production components or limitations leading to these perceptions. Listeners unfamiliar with tracheoesophageal speech listened to normal and tracheoesophageal speakers saying the sentence *Put the dishes away* as a declarative, query, and exclamation—both in and out of sentential context. Listeners then heard a recording of all speakers and categorized the sentences as to sentence types. Intonation patterns of normal and tracheoesophageal speakers were then compared and contrasted. A more detailed analysis of tracheoesophageal tokens identified as correct nearly all or none of the time was also done.

4pSC19. A high-speed methodology for comparing phonosurgical outcomes in an excised larynx model. Niro Tayama (Dept. of Otolaryngol., Head & Neck Surgery, Univ. of Tokyo, 3-7-1, Hongo, Bunkyo-ku, Tokyo 113-8655, Japan), Douglas W. Montequin (Univ. of Iowa, Iowa City, IA), and David A. Berry (UCLA School of Medicine, Los Angeles, CA 90095-1794)

Many phonosurgical techniques have been used to treat hoarseness resulting from vocal disease, such as unilateral vocal fold paralysis. However, because of difficulty in documenting changes in mucosal wave propagation, comparing the outcomes of different phonosurgical techniques presents a formidable challenge. The purpose of this investigation is to compare two phonosurgical treatments of unilateral vocal fold paralysis, thyroplasty type 1 and arytenoid adduction, in terms of their impact on the mucosal wave. For thyroplasty type 1, the pushing force was systematically varied. For arytenoid adduction, the pulling forces were systematically varied. Across all experiments, systematic changes were also made in the subglottal pressure, glottal airflow, pre-phonatory posturing, and in the size of the glottal chink. Because stroboscopic methods could not image properly the aperiodic vibrations associated with paralysis, a digital high-speed imaging system was utilized (the Kodak EktaPro 4540). Using this system, the 3-D vibrations of the mucosal wave were imaged across all experimental conditions. In this presentation, we focus on the methodology employed to collect these measurements.

4pSC20. Analysis of acoustic factors contributing to poor quality of electrolaryngeal speech. Geoffrey S. Meltzner (Speech and Hearing Sci. Prog., MIT, Cambridge, MA 02319 and Voice & Speech Lab, Mass. Eye and Ear Infirmary, Boston, MA 02114, geoff@mit.edu), Robert E. Hillman (Harvard Med. School, Boston, MA 02115), and Kenneth N. Stevens (MIT, Cambridge, MA 02139)

A large percentage of patients who have undergone laryngectomy to treat advanced laryngeal cancer rely on an electrolarynx (EL) to communicate orally. While serviceable, EL speech is plagued by shortcomings in both quality and intelligibility. As part of a larger ongoing effort to improve the oral communication of EL users, this study seeks to better quantify the sources and impacts of abnormal acoustic properties typically found in EL speech. Two EL characteristics, altered voicing source location and directly radiated self-noise, were investigated by recording EL speech tokens in two recording conditions: (1) in an acoustic chamber (self-noise present) and (2) through a specially constructed port into an acoustic chamber (self-noise absent). An analysis of several vowels reveals that one or more of the first three formants are often masked by the direct noise and that their amplitudes are affected by EL location on the neck. Results are discussed relative to a comprehensive comparative analysis of pre-laryngectomy versus post-laryngectomy EL speech, and with respect to a model of EL speech production. [Work supported by W. M. Keck Foundation and VA Grant C2243-2DC.]

THURSDAY AFTERNOON, 6 DECEMBER 2001 ROOM 223, 1:30 TO 4:15 P.M.

Session 4pSP

Signal Processing in Acoustics: General Topics in Signal Processing II

Deborah M. Grove, Chair
CAIP Center, Rutgers University, 96 Frelinghuysen Road, Core Building, Piscataway, New Jersey 08854-8088

Contributed Papers

1:30

4pSP1. Statistical control of ultrasonic power measurements using hypothesis testing. Lixue Wu and George Wong (Acoust. Standards, Inst. for Natl. Measurement Standards, Natl. Res. Council Canada, Ottawa, ON K1A 0R6, Canada)

The measurement in the calibration of ultrasonic power was controlled using hypothesis testing. The ultrasonic power measurement system is based on the radiation force balance method according to IEC 61161 (1998). To assess the uncertainty in the measurement, the experimental variance needs to be estimated. Having a well-characterized measurement with statistical method in accordance with the Guide to the Expression of Uncertainty in Measurement (4.2.4 and the note of H.3.6), a pooled estimate of the experimental variance that characterizes the measurement is obtained. During an acoustical calibration and measurements over a period of 6 weeks, hypothesis testing was applied to the measurement data to ensure the validity of the pooled experimental variance. The measurement data were divided into data blocks with a sliding window to form five independent consecutive observations. Hypothesis tests were then applied to the data blocks to examine the variability of the value of a measurand. Good measurement reproducibility was observed that resulted in the confidence of the measurement.

1:45

4pSP2. The effects of fluctuation based processing on waterfall displays. Dale B. Dalrymple (Signal Processing Systems, Smiths Industries Aerosp., 13112 Evening Ck. Dr., San Diego, CA 92128-4199)

Fluctuation based processing is a nonlinear method of processing time series data. Passive sonar waterfall displays are time-frequency plots where sequential data in each bin is a time series. Fluctuation based processing enhances the detection of low-fluctuation amplitude signals when processed against a background of high fluctuation noise or interferers. This paper examines the impact of fluctuation based processing on the signal processing load required for the generation of waterfall displays and examines alternatives to traditional waterfall normalization methods that apply to low-fluctuation amplitude signals.

2:00

4pSP3. Autocorrelation estimation using a traveling-wave filterbank. Timothy A. Wilson (Embry-Riddle Aeronautical Univ., Daytona Beach, FL 32114)

A traveling-wave filterbank was implemented from a causal linear-phase unimodal-magnitude prototype by a time-scale operation similar to the design of a wavelet filterbank. The channels of the resulting filterbank are delayed relative to each other in a monotonic fashion, creating the

traveling-wave effect. The scale varied in an exponential manner with channel number (place). Outputs at two places were multiplied by each other. Were the filterbank simply a nonuniform delay line and the input a stationary random process, each such product would be an unbiased estimator of the input autocorrelation at a lag corresponding the difference between the delays (relative to the input) of each channel. Considering the indices of two channels as coordinates in a Cartesian system, loci of constant relative delay were identified. Products of outputs along those loci were averaged to estimate the autocorrelation at the lag corresponding to that relative delay. Lags were chosen to be exponentially distributed over a certain range, so the collection of estimates form a log–lag autocorrelation. Estimates and their performance are compared to those of correlogram-based methods.

2:15

4pSP4. CW-based proximity distance measurement by ultrasonic sensors. Xi Li, Jian Li, and Mark Sheplak (Dept. of Elec. and Computer Eng., P.O. Box 116130, Univ. of Florida, Gainesville, FL 32611, xili@dsp.ufl.edu)

With conventional pulse-echo ranging approaches, it is difficult to obtain an accurate range measurement needed by a proximity ultrasonic sensor system due to the limited bandwidth of the transducer. However, based on the continuous wave (CW) method (i.e., the transducer is excited by a set of continuous harmonic signals), the high accuracy can be obtained even with a narrow-band transducer. By using a set of orthogonal sinusoid signals and simultaneously measuring the phase shifts corresponding to all of the frequencies, we obtain a fast measurement scheme with respect to the traditional CW method that measures the phase shifts for each frequency separately. This method can be efficiently implemented by FFT by currently available dedicated DSP chips. To estimate the unknown distance based on the phase shifts, two-stage fast optimization algorithm has been presented. Numerical results show that this algorithm can approach the Cramer–Rao bound (CRB) for high signal-to-noise ratio.

2:30

4pSP5. Recent advances in structural processing: Resolution of short transients of unknown parameters. Eugene I. Plotkin and M. N. S. Swamy (Concordia Univ., Ctr. for Signal Processing and Commun., 1455 De Maisonneuve Blvd., Montreal, PQ H3G 1M8, Canada)

Resolution of transients of identical structures arises in such areas as machinery vibration, seismic and underwater acoustics. These processes are characterized by highly correlated nonstationary nature over a short interval. In extreme case, where the transients have the same structure and the record-length is short, even the most powerful adaptive methods become inefficient. Composite parameter-free modeling (CPFM) [E. I. Plotkin and M. N. S. Swamy, Int. J. Acoust. Vib. **4**, 159–164 (1999)] is one of the effective techniques for improving the resolution of closely spaced (in time and frequency) short transients of unknown parameters. The model presented is a nested-form composition of null filters; the inner building blocks (variable-frame matched filters) are used to suppress one of the transients, while the outer nonlinear structure nullifies the second transient. The proposed model permits linear estimation of the target transient corrupted by almost identical targetlike interference. This model exhibits clear advantage in reconstructing a target signal in the presence of a powerful multi-tone transient of unknown parameters. The estimated envelope and phase of the target rapidly converge to their steady-state values, while the conventional approach produces prolonged lags with extensive fluctuations, precluding reliable reconstruction of the target signal. [Work supported by the NSERC.]

2:45–3:00 Break

3:00

4pSP6. The matched-lag filter: Detecting broadband multipath signals with auto- and cross-correlation functions. John L. Spiesberger (Dept. of Earth and Environ. Sci., Univ. of Pennsylvania, 240 S. 33rd St., Philadelphia, PA 19104-6316, johnsr@sas.upenn.edu)

Signal detection is considered for uncertain noise variance and a broadband source of unknown waveform and emission time. Multipath travel times may be unknown. Using a new ''matched-lag filter,'' the presence or absence of the signal is estimated from the auto- and cross-correlation functions of the receptions. Like a matched filter, correlation functions provide the first stage of gain in signal to noise ratio because the paths are assumed to be partially coherent. The second stage achieves additional gain by searching only over physically possible arrangements of signals in the auto- and cross-correlation functions while excluding forbidden arrangements. These stages enable the matched-lag filter to behave like a matched filter within a matched filter. In an ideal case, simulations of the matched-lag filter yield probabilities of detection that are, with one and two receivers, 4.1 to 366 times respectively that obtained from the conventional energy detector at a false-alarm probability of 0.001. The matched-lag filter has applications to wireless communications and the detection of acoustic signals from animals, vehicles, ships, and nuclear blasts. The matched-lag filter more completely describes signal structure than stochastic detection and communication theories whose specified autocorrelation function does not prohibit forbidden arrangements. [Work supported by ONR.]

3:15

4pSP7. Multi-wavelet detection and denoising of low-frequency chirp signals using adaptive wavelet methods. Joseph S. Wheatley, Juliette W. Ioup, and George E. Ioup (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, wheatley@datastar.net)

The detection and classification of underwater acoustic signals embedded in noise is a fundamental problem of interest to the signal processing community. The use of wavelet transforms is a recent development in digital signal processing which has been applied in many different areas. A particular type of wavelet is the chirplet, which includes frequency variation as well as time shift and scaling. Both linear and polynomial chirp signals are present in underwater acoustic signals generated by such sources as biologics, ships, and submarines. Distinguishing the features of these chirps relative to other ambient noise shows promise as an initial step in classification of underwater acoustic signals. Removal of unwanted broadband signal components via wavelet methods has been shown to outperform other noise removal processes such as low-pass and high-pass filtering and Weiner filtering. Examples of low-frequency simulated chirp signals with additive noise have been generated. A multi-wavelet packet method for detection and denoising low-frequency signals containing multiple chirps embedded in noise using a specific wavelet designed for polynomial chirp signals is shown. [Research supported in part by ONR.]

3:30

4pSP8. Short-time fractional Fourier methods for signal representation in the time-frequency plane. Chris Capus and Keith Brown (Image Anal. Res. Group, Dept. of Computer and Elec. Eng., Heriot-Watt Univ., Riccarton, Edinburgh EH14 4AS, UK, ceecgc@cee.hw.ac.uk)

The analysis of multicomponent chirp signals is of interest in sonar, geophysical, bioacoustic, ultrasonic, and radar applications. This paper examines the interpretation of the fractional Fourier transform (FrFT) of such signals in terms of their representation in the time-frequency plane. The FrFT has been shown to be of value in analyzing and filtering linear chirp signals with various characteristics. The desirability of a short-time variant has been suggested by several authors and this paper introduces

new short-time fractional Fourier methods for the piecewise description and parameterization of nonlinear chirp signals. Understanding of a signal's FrFT as the projection of the joint time-frequency distribution onto a line at angle α to the time axis permits development of these short-time methods. Parameter α provides a means for transform optimization with reference to the entire signal or on a windowed basis. While providing much improved time-frequency resolution over the spectrogram, these techniques do not suffer from cross-term interference seen in Wigner distributions and related methods. Representation of highly nonlinear chirps in the time-frequency plane is demonstrated through analysis of a variety of signals including bat and dolphin chirps, heart sounds, and synthetic signals constructed to highlight particular differences between the techniques.

3:45

4pSP9. Automatic recognition of working states of a rock breaker machine from acoustic signal processing. Juan R. Aguilar, Renato A. Salinas, and Homero E. Latorre (EE Dept., Univ. of Santiago, Ave. Ecuador 3519, Santiago, Chile, rsalinas@lauca.usach.cl)

In underground mining, there are risks involved in operating heavy machinery such as Load–Haul–Dump and rock breakers. In order to minimize those risks, that can be fatal, it was proposed to study the feasibility of employing both vision and sound to improve either human-based teleoperation or robotized machinery. Earlier attempts have considered only visual input data. In this work, recognition results carried out on acoustic emissions generated by a rock breaker during normal operating activities, are presented. These acoustic data were recorded *in-situ* and then analyzed using four different techniques: FFT, FFT-Wavelet, Cepstrum, and Time–Frequency. The results indicate that Power Cepstrum analysis provides the best performance both in terms of finding noticeable differences between the signals representing the operating states of the rock breaker as well as in terms of data compression capabilities. A back-propagation neural network was then trained to classify the signals into the different operating states of the machine. [Work partially supported by Project Dicyt 009913SS, University of Santiago, Chile.]

4:00

4pSP10. Potential application of CLEAN algorithm for ultrasonic NDT. Khan Mohammad Mahmud and Ryoji Ohba (Hokkaido Univ., Faculty of Eng., Div. of Appl. Phys., Kita-13, Nishi-8, Sapporo 060-8628, Japan, mah@eng.hokudai.ac.jp)

The detection of close reflection peaks in ultrasonic pulse compression system has been studied by implementing the CLEAN algorithm, which is popular in image processing of radio astronomy data. The algorithm has been adjusted so that it is applicable to ultrasonic data. It has been applied to some practical data with satisfactory results. It is found to be effective for the detection of close peaks and removing sidelobes. The algorithm has been implemented successfully in the detection of close holes in metal plates and some potential use of the method has been described.

THURSDAY AFTERNOON, 6 DECEMBER 2001 ROOM 304, 1:00 TO 4:45 P.M.

Session 4pUW

Underwater Acoustics: Reverberation in Shallow Water II

Eric I. Thorsos, Chair

Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

Contributed Papers

1:00

4pUW1. Unperturbed normal mode method for surface wave effects on acoustic propagation in a shallow water waveguide. Frank S. Henyey, Eric I. Thorsos, and Stephen A. Reynolds (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, frank@apl.washington.edu)

A computational method is presented for propagation in an ocean waveguide with a wavy surface. This method is based on the expansion of the wave function into the vertical normal modes of the flat-surface problem. Unlike the traditional boundary perturbation technique for elliptic problems, this method is incorporated into an initial-value normal mode framework. The truncated modal expansion is supplemented by an additional function to remove the Gibbs phenomenon at the unperturbed surface. This additional function is approximately a sine integral. The mode coupling equations and the expression for the amplitude of the sine integral are derived. In comparison to local normal modes, the coupling is local at the surface, and the modes are range independent if the unperturbed waveguide is. The coupling matrix is rank 1, enhancing the computational efficiency. [Work supported by ONR.]

1:15

4pUW2. Acoustic propagation in a shallow water waveguide with a rough surface: Comparison of results from PE and coupled unperturbed normal mode methods. Stephen A. Reynolds, Eric I. Thorsos, and Frank S. Henyey (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, steven@apl.washington.edu)

The PE method is an accepted standard for accurate acoustic propagation modeling. We have developed an alternative method (Henyey *et al.*, this session) based on flat-surface normal modes, with coupling due to surface waves. Results for the complex field obtained with the new mode coupling method are compared with the solutions from a rough surface PE code [A. D. Rosenberg, J. Acoust. Soc. Am. **105**, 144–153 (1999)]. A variety of rough surfaces are considered in the comparison. Mode coupling is most easily tested using a single mode starting field with a sinusoidal surface whose wavelength is chosen to produce a resonance with another mode. As an example of a randomly rough surface, a modified Pierson–Moskowitz spectrum of surface waves is considered. Comparisons with the PE method include closeness of the results, accuracy of energy conservation, computational speed, and ease of use. [Work supported by ONR.]

1:30

4pUW3. Rough boundary effects on shallow water propagation: Some comparisons between GRAB and PE. Kevin L. Williams, Eric I. Thorsos, and W. T. Elam (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

The Gaussian Ray Bundle (GRAB) computer code is an accepted Navy Standard for calculating propagation. GRAB uses an effective reflection loss to approximate the effect of ocean surface roughness. Results for GRAB and rough surface PE are compared for a shallow water waveguide (50 m depth with a sand bottom) for several wind speeds and for ranges up to 5 km. A special case examined is a 3.2 kHz source at mid depth with a 10° beamwidth and 10° grazing angle. The GRAB results use the Bechmann-Spezzichino reflection coefficient module. The PE results are calculated using 50 realizations of a Pierson–Moskowitz surface and either coherent or incoherent averaging. GRAB with Bechmann–Spezzichino is most appropriately compared to the PE coherent averages. This comparison shows differences on the order of 3 dB in the intensity averaged across water depth at 5 km and a wind speed of 5 m/s. The general implications of these differences as well as differences seen in the intensity as a function of depth at discussed. Work supported by the ONR.

1:45

4pUW4. Novel physical interpretations of seafloor scattering leading to K-distributed clutter. Anthony P. Lyons and Douglas A. Abraham (The Pennsylvania State Univ., Appl. Res. Lab., P.O. Box 30, State College, PA 16804-0030)

Interest in describing and modeling envelope distributions of seafloor backscatter has increased recently, particularly with regard to high-resolution sonar systems. Seafloor scattering that can be described by heavy-tailed envelope distribution functions (i.e., non-Rayleigh distributions) is often the limiting factor in the performance of these types of sonar systems and in this context is referred to as acoustic clutter analogous to radar clutter. Modeling of clutter has traditionally entailed fitting various candidate distributions to time samples of the envelope of the scattered sonar (or radar) returns. This type of descriptive analysis yields little insight into the environmental mechanisms responsible for heavy-tailed distributions (e.g., distributions and clustering of discrete scatterers, patchiness in geo-acoustic properties, scattering strength of scatterers, etc.). In contrast, we derive the envelope distribution for the scattered returns starting from simple physical descriptions of the environment. It will be shown that several realistic descriptions of the environment lead to K-distributed clutter. This result explains, at least partially, the success of the K-distribution in the modeling of radar clutter and sonar reverberation at a variety of frequencies and scales. These results should prove useful in developing methods for modeling, predicting and mitigating clutter on high-resolution sonar systems. [Work sponsored by ONR under Contract No. N00014-01-1-0352.]

2:00

4pUW5. Parabolic wave equation based rough sea surface scattering. Frank Ryan (CODE D71, SPAWAR Systems Ctr., 53560 Hull St., San Diego, CA 92152-5001, ryan@nosc.mil)

A physics based technique is described for computing both forward scattered and backscattered (i.e., reverberation) fields from realistic rough sea surfaces at cm wavelengths. The method is based on a novel split-step Fourier parabolic wave equation (PWE) method incorporating a stochastic wind and swell ocean wave spectral model which includes both gravity and capillary scales. Comparisons between the PWE based scattered fields and results using the standard Miller–Brown rough surface reflection coefficient method will be shown. Examples of the rough surface scattered field vertical angle-of-arrival spectra under a variety of ducting conditions will be shown, and the impact on high precision spatial tracking (beamforming) discussed.

2:15

4pUW6. Acoustic wave detection and analysis in a sandy sediment. Nicholas P. Chotiros (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029, chotiros@arlut.utexas.edu)

Coherence, wave front curvature, and direction of arrival provide important clues for analyzing acoustic signals in ocean sediments. Measurements were made using a buried acoustic receiving array and a mobile sound projector to study the underlying physical processes in the penetration of sound into sandy ocean sediments, particularly at shallow grazing angles, as part of the Sediment Acoustics Experiment (SAX99). The buried array was insonified by a wide-band sound source carried on a remotely operated vehicle (ROV). The discrimination between refraction and scattering processes was of particular interest. A method was used that employed coherent superposition to distinguish between refracted and scattered signals. Variable focusing was used to sharpen the distinction between refraction and scattering. Sound waves entering the sediment at steep angles were clearly refracted. At shallow angles, other processes come into play. This is a component of a multi-disciplinary effort to study sediment acoustics. [Work supported by ONR, Ocean Acoustics.]

2:30

4pUW7. Determination of the character of sediments and model verification using frequency-dependent reflection coefficients. Marcia J. Isakson, Egbert HoodIII, and Nicholas P. Chotiros (Appl. Res. Labs., The Univ. of Texas, Austin, TX 78713-8029, chotiros@arlut.utexas.edu)

Several models exist for acoustic interaction with ocean sediments, including solid, fluid, and porous media representations, with and without interface roughness and volume scattering mechanisms. We present a method for model verification and determining the contribution from each mechanism by measuring the reflection coefficient. The method is self-normalizing and designed to minimize near-field effects. Data from smooth and roughened surfaces of unwashed river sand will be presented. Measurements were taken using a chirped signal from 50 kHz to 150 kHz at incident angles from 5 to 70 deg. Comparisons are drawn between models and experimental results. [Work supported by ONR, Ocean Acoustics.]

2:45

4pUW8. Centimeter scale seafloor relief measurement using an ROV. Nicholas P. Chotiros, James N. Piper, James Moore, and Daniel F. Weigl (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029, chotiros@arlut.utexas.edu)

For high-frequency acoustic backscatter modeling, in the band 1 to 100 kHz, centimeter scale seafloor roughness is an important parameter. A measurement method was devised for deployment aboard a remotely operated vehicle (ROV) consisting of lasers and a video camera. It was designed to measure small-scale roughness in the range of spatial wavelengths from 0.01 to 1 m. The ROV is well suited to this application because it can hover above the seabed, and it can cover a relatively large area. A cylindrical lens was used to spread the laser beam from a low-power laser into a fan, and direct it downward to produce a stripe on the sediment. Viewing from an oblique angle with a video camera, the profile of the sediment directly under the stripe was obtained. The most difficult problem was precisely determining and tracking the position of the ROV. In the SAX99 experiment, positioning was achieved by using small shells and other objects on the seafloor as landmarks. [Work supported by ONR, Ocean Acoustics.]

3:00–3:15 Break

3:15

4pUW9. Rapid assessment of gently varying bathymetry using TRM reverberation. B. Edward McDonald (U.S. Naval Res. Lab, Washington, DC 20375) and Charles Holland (Penn State Univ., State College, PA 16804)

A concept for rapidly probing shallow water bathymetry [B. E. McDonald and Charles Holland, J. Acoust. Soc. Am. **109**, 2495 (2001)] using reverberation from a time reversed mirror (TRM) is further developed. Adiabatic modal analysis and WKB methods applied to a mildly range-

dependent shallow water waveguide lead to a simple relation between TRM focal shifts and range integrated bathymetric variations. This relation is compared with RAM simulations carried out in conjunction with the July 1999 Saclantcen–Scripps bottom reverberation experiment near Elba. Our integral expression is found to agree satisfactorily with focal spot shifts predicted by RAM. Simulation studies using the Saclantcen PROSIM reverberation model are also presented to help explain details of the July 1999 data. Experimental implications for rapid assessment of unknown or shifting shallow bathymetry are discussed. [Work supported by NRL.]

3:30

4pUW10. Probability of detection and false alarm for target signals measured with seafloor reverberation in a shallow water waveguide. Purnima Ratilal and Nicholas C. Makris (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

Probabilities of detection and false alarm are calculated as a function of the signal-to-diffuse-reverberation ratio for both deterministic and fluctuating targets in a variety of shallow water waveguides. The signal-to-diffuse-reverberation ratio is computed using the unified model [N. C. Makris and P. Ratilal, J. Acoust. Soc. Am. **109**, 909–941 (2001)] for reverberation and submerged object scattering in a stratified medium. The reverberant background is range and azimuth dependent, and varies with the resolution footprint of the sonar system. Scenarios involving an active source and a horizontal receiving array are investigated. Predictions using this method are compared with data from the Geoclutter 2001 experiment. Detection and false alarm probabilities are calculated for submerged targets as well as geoclutter features in the shallow waters of the Stratafrom site off the coast of New Jersey. The results are consistent with our experimental findings that detection probabilities decrease with increasing range.

3:45

4pUW11. Spatially coincident measurements of acoustic backscatter and forward scatter at shallow grazing angles. Paul C. Hines, John C. Osler, Shawn A. Stewart, and Darcy J. MacDougald (Defence Res. Establishment Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada)

Direct measurement of acoustic scattering from the seabed at shallow grazing angles presents a considerable challenge in littoral waters. Specifically, returns from the air–water interface typically contaminate the signal of interest. To address this issue DREA has developed a pair of sea going research systems for measuring acoustic scatter from the seabed in shallow water environs. The first system is a bottom mounted parametric transmitter with a 6-channel superdirective line array receiver. The second system is a vertical line array of 8 hydrophones and 2 transmitters. Operating the systems in tandem allows measurements of backscatter and forward scatter from the same scattering patch at the same time. In this paper the acoustic attributes of both systems will be outlined, the experimental geometry will be described, and measurements of acoustic backscatter and forward scatter collected during one such experiment will be presented.

4:00

4pUW12. Relating changes in reverberation envelope statistics to changes in sonar system parameters. Douglas A. Abraham (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804)

The reverberation component of the matched filter envelope in active sonar systems operating in shallow water environments has often been observed to have probability density function tails heavier than the traditionally assumed Rayleigh distribution. Recent research has shown that exponential point scattering leads to a K-distributed envelope (D. A. Abraham and A. P. Lyons, in Proceedings of the MTS/IEEE Oceans Conference, Honolulu, Hawaii, Nov. 2001). Under this model, the reverberation statistics change as the number of scatterers within a range-bearing resolution cell changes. For example, doubling the bandwidth of a transmission will halve the number of scatterers, resulting in a heavier reverberation distribution tail. This relationship is examined using low frequency active sonar data collected during SACLANT Undersea Research Center's SCARAB 1997 sea-trial (Charles Holland, Chief Scientist) by evaluating the *effective number of scatterers* parameter (α) of the K-distribution as estimated from the data for varying transmit waveform bandwidths and receiving array beamwidths. It was observed that, within the accuracy of the parameter estimation, α varied as the scattering model predicted; that is, it changed proportionally to resolution cell size. [Work sponsored by ONR under Contract No. N00014-00-G-0058, D.O. 0009.]

4:15

4pUW13. A hidden Markov approach for the probabilistic treatment of competing acoustic scattering mechanisms. Raymond J. Soukup and William C. Howell (Naval Res. Lab., Washington DC)

A variety of acoustic scattering phenomena are encountered in the ocean, such as the rough air–sea interface, bubble clouds, biologics, and the ocean bottom. As a result of their relative dependence on sonar parameters such as transmit frequency and launch angle, there are scenarios where a dominant phenomenon (or state) can be observed in reverberation data. Furthermore, important environmental parameters needed by physics-based models to predict such scattering behavior are either temporally/spatially varying or difficult to measure, and must be treated probabilistically. In this paper we present a model based on a hidden Markov approach that describes the probabilities that certain states will be encountered, and that transitions from one state to another will arise due to changes in the environment and the system. This Hidden Markov Model (HMM) integrates physical knowledge of the competing mechanisms in such a way as to illustrate the overall scattering behavior of the ocean environment and to suggest methodologies for reducing the local reverberation. The virtues and features of the HMM are demonstrated using measured data collected in the Gulf of Alaska, where near-surface scattering due to the air-sea interface, bubble clouds, and salmon are all possibly dominant under certain conditions. [Work supported by ONR.]

4:30

4pUW14. The spectra of energy scattered by dispersed anchovies. Orest Diachok (Naval Res. Lab., Washington, DC 20375)

The spectra of backscattered energy by dispersed anchovies, which were reported by Holliday (1972), reveal several peaks at frequencies that correspond to theoretically calculated resonance frequencies of year classes of anchovies. Theoretical calculations are based on concurrent measurements of distributions of swim bladder dimensions and a modified form of Minnaert's equation. Differences between calculated and measured values of the mean lengths of the second, third, and fourth year classes are within experimental uncertainties (8%). The calculated mean lengths of juvenile anchovies are in good agreement with historical measurements of this parameter. Matching of calculations and measurements of backscattered energy level versus frequency yields estimates of the total Q of spectral lines and the relative number density per year class. The resultant estimate of Q of adult anchovies is approximately 4.4. This value of Q is consistent with laboratory measurements of the Q of individual anchovies ($\sim$7 at 15 m) and measurements of length and depth distributions of year classes. Resultant estimates of relative number densities of year-classes were consistent with historical measurements of this parameter in the Southern California Bight. [Work supported by ONR.]

Meeting of Accredited Standards Committee (ASC) S1 Acoustics

G. S. K. Wong, Chair S1
Institute for National Measurement Standards, National Research Council, Montreal Road, Bldg. M36, Ottawa, Ontario K1A OR6, Canada

T. F. W. Embleton, Interim Vice Chair S1
80 Sheardown Drive, Nobleton, Ontario LOG 1N0, Canada

P. D. Schomer, Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 43, Acoustics
2117 Robert Drive, Champaign, Illinois 61821

H. E. von Gierke, Vice Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 43, Acoustics
1325 Meadow Lane, Yellow Springs, Ohio 45387

V. Nedzelnitsky, U.S. Technical Advisor (TA) for IEC/TC 29, Electroacoustics
National Institute of Standards and Technology (NIST), Sound Building, Room A147, 100 Bureau Drive, Stop 8221, Gaithersburg, Maryland 20899-8221

Accredited Standards Committee S1 on Acoustics. Working group chairs will report on the status of standards currently under development in the areas of physical acoustics, electroacoustics, sonics, ultrasonics, and underwater sound, etc. Consideration will be given to new standards that might be needed over the next few years. There will be a report on the interface of S1 activities with those of ISO/TC 43 Acoustics and IEC/TC 29 Electroacoustics, including plans for future meetings of these Technical Committees. The Technical Advisory Groups for these Technical Committees consist of members of S1, S3, and other persons not necessarily members of those Committees. Open discussion of committee reports is encouraged.

Scope of S1: Standards, specifications, methods of measurement and test, and terminology in the field of physical acoustics, including architectural acoustics, electroacoustics, sonics and ultrasonics, and underwater sound, but excluding those aspects which pertain to biological safety, tolerance and comfort.

Meeting of Accredited Standards Committee (ASC) S3 Bioacoustics

R. F. Burkard, Chair S3
Hearing Research Laboratory, State University of New York at Buffalo, 215 Parker Hall, Buffalo, New York 14214

J. Franks, Vice Chair S3
Robert A. Taft Laboratories, 4676 Columbia Parkway, Mail Stop C27, Cincinnati, Ohio 45226

P. D. Schomer, Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 43, Acoustics
2117 Robert Drive, Champaign, Illinois 61821

D. D. Reynolds, Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 4, Human Exposure to Mechanical Vibration and Shock
3939 Briar Crest Court, Las Vegas, Nevada 89120

H. E. von Gierke, Vice Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 43, Acoustics and ISO/TC 108/SC 4, Human Exposure to Mechanical Vibration and Shock
1325 Meadow Lane, Yellow Springs, Ohio 45387

Accredited Standards Committee S3 on Bioacoustics. Working group chairs will report on the status of standards under development. In addition to those topics of interest, including hearing conservation, noise, dosimeters, hearing aids, etc., consideration will be given to new standards that might be needed over the next few years. There will be a report on the interface of S3 activities with those of ISO/TC 108/SC4 Human exposure to mechanical vibration and shock, ISO/TC43 Acoustics, and IEC/TC 29 Electroacoustics, including plans for future meetings of these Technical Committees. The US Technical Advisory Groups for these Technical Committees consist of members of S3, S1 and other persons not necessarily members of those Committees. Open discussion of committee reports is encouraged.

Scope of S3: Standards, specifications, methods of measurement and test, and terminology in the fields of psychological and physiological acoustics, including aspects of general acoustics, shock, and vibration which pertain to biological safety, tolerance, and comfort.

Session 5aAB

Animal Bioacoustics: Session Honoring William Watkins

Peter L. Tyack, Chair
Department of Biology, Woods Hole Oceanographic Institution, 45 Water Street, Woods Hole, Massachusetts 02543

Chair's Introduction—8:30

Invited Papers

8:35

5aAB1. The development and use of an implantable radio whale tag from 1970 to the present. Douglas Wartzok (Florida Intl. Univ., University Park PC236, Miami, FL 33199) and Romaine R. Maiefski (Ontogen Corp., Carlsbad, CA 92009)

William A. Watkins, Woods Hole Oceanographic Institution (WHOI), led the development and refinement of the WHOI Radio Whale Tag. The initial tags were developed in the 1960s at WHOI, improved in the 1970s and 1980s by Ocean Applied Research, Inc., and adapted to satellite telemetry in the 1990s by Microwave Telemetry, Inc. Through this entire period, the tag delivery and attachment methods were continually refined. The same delivery and attachment methods were used with acoustic tags, radio tags, combined radio and acoustic tags, and satellite tags. Scientific studies using these tags have been conducted on finback, humpback, Brydes, bowhead, and sperm whales. These field studies have ranged from Venezuela to the Canadian arctic and Alaska, and from the Pacific to the Mediterranean. Every tagging experiment led to new information about the animals, and in most cases the information from radio and acoustic tagged whales showed that the best available data based on surface observations of untagged animals were deficient. Highlights from some of these findings will be presented.

8:55

5aAB2. Bill Watkins and pinniped bioacoustics. Ray G. Carleton (Dept. of Environ. Sci., Univ. of Virginia, 291 McCormick Rd., Box 400123, Charlottesville, VA 22904)

Bill's work on pinnipeds began with a 1963 paper by Schevill, Watkins, and myself in Science on the underwater sounds of pinnipeds, which resulted from his recordings in the highly unnatural waters of the New York Aquarium. It was written partly in reaction to assumptions being made that seals, like cetaceans, possessed sonar. Following that mode, there followed a paper by the same team in the same location on the remarkable bell tones and knockings of captive walruses. Recording from natural situations, with Schevill in McMurdo Sound, Southern Ocean, resulted in the first analysis of Weddell seal sounds. In 1969, he provided the first technical description of a marine mammal song (in the behavioral meaning of that term), that of the bearded seal, and in 1975 a detailed description of walrus sounds in association with their reproductive behavior. Sounds of other polar pinnipeds, ribbon seals of the Arctic and Ross seals of the Southern Ocean, bring the history up to date. Pinnipeds possess a suite of underwater sounds that rivals any other group. They make up for apparent lack of echo-location ability by musical quality of voice.

Contributed Paper

9:15

5aAB3. High-frequency harmonics and source level of humpback whale songs. Whitlow Au (Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734), Darling James (West Coast Whale Res. Foundation, Vancouver, BC V6R 2L3, Canada), and Kimberly Andrews (Hawaii Inst. of Marine Biol., Kailua, HI 96734)

The songs of 8 male humpback whales were recorded at ranges varying from 20 to 40 m with a calibrated system that had a flat frequency response to 22 kHz. The songs consisted of bursts of sounds called units. Units were organized into phrases and phrases into themes. Most of the units had mean duration between 1 and 2 s and mean silent periods between units between 1 and 2 s. Many of the recorded songs contained units that had high-frequency harmonics that extended beyond 15 kHz. The amplitudes of the higher-frequency harmonics of some units were within 18–24 dB of the fundamental or the highest level harmonic at frequencies out to 13.5 kHz. These harmonic results suggest that humpback whale songs have a broadband quality not previously reported and may provide some insights on the high-frequency limit of hearing in these whales. The source levels of the different songs were also estimated by considering the root mean square sound pressure level referenced to 1 m for the unit with the largest level for each different phrase within a song. Source levels varied between 171 and 189 dB re: 1 μPa.

Invited Paper

9:30

5aAB4. Bill Watkins listens to all the whales in the North Pacific Ocean. Robert Gisiner (Office of Naval Res., 800 N. Quincy St., Arlington, VA 22217-5660)

In the early 1990s portions of the Navy's classified underwater listening systems (IUSS) were made available for research. Bill Watkins had long been familiar with this system and eagerly went to work listening to whales in the North Pacific Ocean. A decade later the work that Bill Watkins has amassed, and its impact on marine mammal research, is truly remarkable. Each month Bill's team

produces a summary of hundreds, if not thousands, of whale calls detected and localized using the Navy's listening arrays. These data have pinpointed seasonal centers of vocal activity in unexpected places, leading to changes in our strategies for monitoring several populations of endangered large whales. Bill has tracked the migratory paths of blue, fin, and humpback whales and has charted the shifts in migratory timing caused by climatic phenomena such as ENSO. Bill has taken a complex technology created for other purposes and crafted a process to get it to deliver simple, clear, and unclassified information for an entire ocean: we have not yet plumbed the limits of what we can do with this information.

9:50–10:05 Break

Contributed Papers

10:05

5aAB5. Multiyear study of blue whale calls in the north Pacific. John A. Hildebrand, Mark A. McDonald (Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0205), Rex Andrew, James A. Mercer, and Bruce M. Howe (Appl. Phys. Lab, Univ. of Washington, Seattle, WA 98105-6698)

Multiyear acoustic recordings provide new insight on blue whale behavior in the Northern Pacific. We analyze six years of blue whale acoustic recordings made in the North Pacific as part of the ATOC (Acoustic Thermometry of Ocean Climate) and NPAL (North Pacific Acoustic Laboratory) projects. Blue whale calls are recorded as a component of ambient sound data, preserved as spectra with 500 Hz Nyquist frequency in 1 Hz bins, averaged over 170 s and recorded at 5 min intervals at each of 13 sites from 1994 to the present. These data verify a separation of blue whale calls into a Western Pacific version with two primary frequencies near 20 Hz and weak harmonic components, and an Eastern Pacific version with a primary near 17 Hz, a strong harmonic component near 50 Hz and a strong overtone near 90 Hz. The frequency of the Eastern Pacific blue whale call has shifted downward throughout the last 40 years, and the present data, combined with published sources, show the frequency shift to be secular and nearly linear with time. [Work supported by ONR.]

10:20

5aAB6. Dive patterns and source levels of sperm whales off New Zealand. Adam S. Frankel (Marine Acoust., Inc., 901 N. Stuart St., Ste. 708, Arlington, VA 22203)

Diving sperm whales off Kaikoura, NZ were recorded using a 4-element vertical linear hydrophone array deployed from a small boat. The boat was frequently within close proximity to the whale dove. Whales began to click almost immediately after leaving the surface. Dive patterns of whales were determined by estimating the location of each click using an advanced time-of-arrival algorithm developed at Cornell University. Analysis of dive patterns showed near vertical descents with speeds ranging from 1–2 m/s. Received click levels were determined and source levels were estimated using a spherical-spreading loss term, as the ranges to the whales were less than the water depth. Preliminary estimates of peak-to-peak source levels range from 191 to 201 dB re: 1 Pa at 1 m (peak-to-peak). On one occasion, a sperm whale clicked rapidly at close range ($\sim$8 m) from a hydrophone element. Click rates as high as 22/s suggest that the whale was interrogating the metal weight at the end of the array. The source level of these clicks ranged from 170 to 205 dB. [Research supported by the Smithsonian Institute and the Cornell University Bioacoustics Research Program.]

10:35

5aAB7. Blue whale call intensity varies with ambient noise level. Sean M. Wiggins, Erin M. Oleson, and John A. Hildebrand (Scripps Inst. of Oceanogr., Univ. of California, San Diego, 9500 Gilman Dr., #0205, La Jolla, CA 92093, swiggins@ucsd.edu)

Blue whales (*Balaenoptera musculus*) vary the intensity of their sound production level in response to varying ambient noise levels. Blue whales produce tonal, type B calls with a fundamental frequency of 16 Hz, and a strong third harmonic at 48 Hz. These calls, along with other low-frequency baleen whale vocalizations, ambient sea noise, and shipping noise have been monitored and recorded on continuous seafloor acoustic recorders near Cortez and Tanner Banks, offshore Southern California, since August 2000. By separately comparing the intensity in the 16 and 48 Hz spectral level bands with the intensity of ambient noise at and near those frequencies, we observe a correlation between the sound intensity level of blue whale type B calls and ambient noise levels. This is verified by the inspection of individual call intensity in various noise environments. Preliminary analysis indicates a less than unity proportional increase in the call intensity with increasing ambient noise. A trend of increased call intensity with an increased noise level has implications for the effect of shipping, and other natural and artificial noise sources in the marine environment on the behavior of blue whales.

10:50–11:20
Closing Remarks

Session 5aSA

Structural Acoustics and Vibration: Shells and Plates

Jeffrey S. Vipperman, Cochair
Department of Mechanical Engineering, University of Pittsburgh, 648 Bennedum Hall, 3700 Ohara, Pittsburgh, Pennsylvania 16804

Courtney B. Burroughs, Cochair
Applied Research Laboratory, Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

Contributed Papers

8:30

5aSA1. How to label resonances from elastic shells. H. Uberall (Phys. Dept., Catholic Univ. of America) and M. F. Werby (NRL Code 7181, Stennis Space Center, MS)

A large class of resonances is observed from the excitation of elastic shells. A review of the type of resonances is presented. It is then shown how to label the resonances depending on material properties, ambient fluid, shape of object, and evacuation or fluid filling.

8:45

5aSA2. Tranmission loss of a ribbed composite vessel. Jeffrey S. Vipperman, Deyu Li, and Ilya Avdeev (Dept. of Mech. Eng., Univ. of Pittsburgh, 531 Benedum Hall, Pittsburgh, PA 15228)

The tranmission loss behavior of a new type of composite structure has been characterized through experimental structural and acoustic modal analyses as well as finite element analysis (FEA) modeling. The tapered cylindrical structure, which is being considered for expendable launch vehicles, is fabricated graphite-epoxy and has a series of internal tridirectional rib stiffeners. Light coupling between the structure and interior acoustic cavity exists, permitting separate analyses of the ''structurally dominant'' and ''acoustically dominant'' modes for the system. The modal analyses and FEA modeling efforts are complimentary since the former helps to determine appropriate FEA model parameters and the latter predicts the shape and ordering of the modes to expect in the experiments. These modes can be subsequently coupled through the modal-interaction approach for prediction purposes. In addition, they can be used to determine the structural acoustic coupling coefficients or radiation modes of the system in order to assess the relative contributions of various structural modes to the internal acoustic levels. A comparison between *in situ* transmission loss measurements and the prediction capabilities of the FEA models will be assessed. [Work supported by AFRL Space Vehicles Directorate, and University of Pittsburgh.]

9:00

5aSA3. Piezoelectric sensors on internally fluid loaded elastic shells. Peter R. Stepanishen (Dept. of Oceanic Eng., Univ. of Rhode Island, Narragansett, RI 02882) and Donald L. Cox (Naval Underwater Warfare Ctr. Div. Newport, Newport, RI)

A general analytical approach is presented to evaluate the harmonic response of electro-mechanical–acoustical systems consisting of external piezoelectric sensors on internally fluid loaded elastic shells with electrical, mechanical and/or acoustical excitations. The approach utilizes the *in-vacuo* eigenvectors of the shell as the basis functions for the solution of the fluid loaded shell and sensor interaction problem. Fluid loading on the shell is included in the approach via internal mode and frequency-dependent radiation impedances and multiport models are used to represent the sensors. As a specific case of interest, the response of conformal piezoelectric sensors on an internally fluid loaded spherical shell is addressed using both the analytical approach and the FEM method. Numerical results are presented to illustrate the importance of acoustic and elastic interaction effects among sensors on the shell with and without internal fluid. Agreement between the FEM and analytical results is excellent over a broad frequency range. General characteristics of the electro-mechanical–acoustical responses are discussed. [Work supported by ONR.]

9:15

5aSA4. Backward projection of harmonic pressure measurements using the generalized internal source density method and SVD. Peter R. Stepanishen and Irsan S. Brodjoneboro (Dept. of Oceanic Eng., Univ. of Rhode Island, Narragansett, RI 02882)

The Generalized Internal Source Density (GISD) method [P. R. Stepanishen, J. Acoust. Soc. Am. **102**, 1955–1963 (1999)] is used to backward project acoustic harmonic pressure fields from an external measurement surface of revolution enclosing the finite three-dimensional vibrating body that generated the field. Both the normal velocity and pressure fields on the vibrating surface can be readily reconstructed using the method. The GISD method is based on first decomposing the acoustic field on the measurement surface into a modal summation of circumferential orders, where each modal pressure field is associated with an internal linear distribution of ring sources on the axis of revolution of the measurement surface. Each modal linear distribution of ring sources is evaluated using Singular Value Decomposition (SVD) methods. A brief review of the method is presented and discussed and then numerical results are presented to illustrate the accuracy of the method for a spherical shell and finite cylindrical shells with hemispherical end caps. The results show excellent agreement with the prescribed normal velocities and associated pressure fields over a wide frequency range.

9:30

5aSA5. Vibration of *in vacuo* hemiprolate spheroidal shells. Sabih I. Hayek (Dept. of Eng. Sci. and Mech., Pennsylvania State Univ., University Park, PA 16802) and Jeffrey E. Boisvert (Naval Undersea Warfare Ctr. Div. Newport, Newport, RI 02841)

The equations of motion for nonaxisymmetric vibration of hemiprolate spheroidal shells of constant thickness were derived using Hamilton's principle. The thin shell theory used in this derivation includes shear deformations and rotatory inertias. The shell is clamped at the equator and is excited by mechanical surface force fields. The displacements and rotations were expanded in infinite series of comparison functions. These include associated Legendre functions in terms of the prolate spheroidal angular coordinate and circular functions in the azimuthal angle coordinate. Five-branched frequency spectra were computed for several shell thickness-to-length ratios ranging from 0.005 to 0.1, and for various diameter-to-length ratios, including the limiting case of a spherical shell. Numerical results were obtained for the frequency spectra of the driving

and transfer mobilities of these shells due to surface force excitations. Some comparisons of the dynamic response of hemiprolate and full prolate spheroidal shells are presented. [Work supported by ONR and the Navy/ASEE Summer Faculty Program.]

9:45

5aSA6. Low-frequency resonances and scattering enhancements of water-filled cylindrical shells open at each end. Curtis F. Osterhoudt and Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

At low frequencies, the contents and termination of cylindrical steel shells can significantly affect the scattering properties. We have observed a giant resonance associated with the limiting case of a water-filled cylindrical shell open at both ends. The resonance, which is analogous to the fundamental mode of an open organ pipe, occurs because of the strong low-frequency reflection of internal acoustic waves from the open ends of the shell [H. Levine and J. Schwinger, Phys. Rev. **73**, 383–406 (1948)]. Since the wavelength is approximately twice the length L of the cylinder, the wave number radius product ka may be much less than unity when L/a is large. The mode was excited by an external acoustic transducer and was detected either as a backscattering enhancement or (in the case of a sufficiently large cylinder) with a hydrophone within the cylinder. By tuning the frequency of an incident tone burst to the mode frequency, the echo is like the response of a high-Q harmonic oscillator, showing a gradual buildup and decay. [Work supported by ONR.]

10:00–10:15 Break

10:15

5aSA7. Impedance tuning of an electromechanical acoustic liner. Stephen B. Horowitz, Toshikazu Nishida (Dept. of Elec. and Computational Eng., Univ. of Florida, 460 NEB Bldg. 33, Gainesville, FL 32611-6130, steveh23@ufl.edu), Louis N. Cattafesta, III, and Mark Sheplak (Dept. of Aerosp. Eng., Mech., and Eng. Sci., Univ. of Florida, 231 Aero Bldg., Gainesville, FL 32611-6250)

The results of a theoretical and experimental study of the impedance tuning capabilities of an electromechanical acoustic liner are presented. An electromechanical acoustic liner utilizes an array of Helmholtz resonators, each containing an electroactive, compliant-walled cavity, to couple the electrical, mechanical and acoustical energy domains. Due to this coupling, the acoustic impedance of this liner is a function of elements from all of these energy domains. Therefore, the acoustic impedance of the liner can be modified via an electrical network. In this study, a prototype of a single cell of an electromechanical acoustic liner is developed and tested. The prototype consists of a cylindrical Helmholtz resonator with a circular, piezoelectric composite backplate, consisting of a thin brass disk bonded to a disk of PZT. An electrical network of resistors, inductors, and capacitors was connected across the terminals of the PZT to implement tuning. Normal-incidence acoustic impedance measurements were taken in an impedance tube from 1 to 6.4 kHz for various values and combinations of electrical components. The measured impedance spectra are then compared to theoretical impedance spectra based on lumped element modeling.

10:30

5aSA8. Dynamic stability of a beam excited by moving masses. Seroj Mackertich (The Pennsylvania State Univ., Harrisburg, Middletown, PA 17057, oct@psu.edu)

The dynamic stability of an elastically supported Timoshenko beam excited by constant velocity equally spaced traveling masses has been investigated. The regions of dynamic stability are determined for different values of the elastic foundation stiffness. Floquet theory is utilized to study the parametric regions of stability and instability, which are displayed in graphical form. Since the occurrence of this dynamic instability reduces the axial buckling load of the beam, the result is important for the study of buckling of a continuous beam.

10:45

5aSA9. Variable structure control of vibration of the simply-supported plates. Jiaqiang Pan (Dept. of Instrument Sci. and Eng., Zhejiang Univ., Hangzhou 310027, PROC) and Shuwen Pan (Dept. of Civil and Environ. Eng., Univ. of California, Irvine, CA 92697)

Robust active vibration control technology is still on its way to develop, though there has been a great number of studies on active vibration control (AVC) topic published since 1980s. The key is to find a good control strategy fitting AVC use and with strong control robustness, in addition to other qualities. As a newer member of robust control strategies, variable structure control (VSC) has attracted control researchers to study its application potentiality at different engineering areas for recent years. Few of them, however, have already paid attention to use in AVC area. In this paper, VSC design methodology used for AVC of a simply-support rectangular plate is investigated. The control design process at the modal state-space is deduced and the matching conditions introduced. Numerically simulated results by using VSC under scheme of the independent modal space control are obtained and compared to those obtained by using least quadratic control and the poles allocation method. It shows that VSC has advantages over the traditional control strategies at control robustness and can control more vibratory modes. [Work supported by the Zhejiang Science and Technology plan fund, PRC.]

11:00

5aSA10. Hydrofoil singing on fast ferries. Konstantin I. Matveev (Caltech, MC 301-46, Pasadena, CA 91125, matveev@caltech.edu)

Hydrofoil singing phenomenon is sometimes observed on new and repaired fast ferries with hydrofoil systems. Hydrofoil singing is a tonal sound caused by interaction of the foil trailing edge with incoming flow. The intensity of the excited tone often dominates the total noise. Hydrofoil singing has very negative effects: first, the noise level is unacceptable for passenger transportation; second, vibration of the foil structure participating in singing leads to fatigue of the system and crack formation on foils and struts. The problem of hydrofoil singing first emerged on ferries with steel foils, but especially critically it was manifested on hydrofoil systems made from titanium. Experimental results obtained on a recent sea-going hydrofoil vessel are reported. Based on the structural peculiarities of the foil system, vortex shedding process and vibration of the trailing edge are discussed, and a lift-oscillator model is used for mathematical modeling. A new method of the special trailing edge profiling is considered; it allows elimination of the tone without degrading the hydrodynamic characteristics of the high performance vehicles.

11:15

5aSA11. Application of feedback control design to improve the vibrator distortion. Bai-Tang Lee and Yun-Hui Liu (Bldg. 16, 321 Kuang Fu Rd., Sec. 2, Hsinchu, Taiwan 300, ROC)

In the beginning, the power amplifier with vibrator combination of the low frequency primary vibration calibration system has total harmonic distortion (THD) of acceleration higher than 2% at most frequencies. To get better calibrated values and to meet the requirements of the International Organization for Standardization (ISO) 16063-11; a method has been developed to reduce the THD. The feedback control method is designed to reduce the signal distortion from the vibrator. The vibrator signal is measured by an accelerometer located on a moving platform that is driven by the vibrator. This signal is fed into a controller at which the difference between the signal and a pure sine wave is calculated and a modified waveform is obtained. After signal processing on the controller, this modified signal is fed back as the input signal of the vibrator to reduce the THD. As expected, the THD measured is below 2%.

Session 5aSC

Speech Communication: Speech Recognition and Synthesis (Poster Session)

Astrid Schmidt-Nielsen, Chair
Naval Research Laboratory, Code 5513, 4555 Overlook Avenue, SW, Washington, D.C. 20375-5337

Contributed Papers

All posters will be on display from 8:30 a.m. to 11:00 a.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 9:45 a.m. and contributors of even-numbered papers will be at their posters from 9:45 a.m. to 11:00 a.m.

5aSC1. Speaker identification with dynamic synapses. Alireza Afshordi Dibazar, Hassan Heidari Namarvar, Jim-Shih Liaw, and Theodore Berger (Dept. of Biomed. Eng., Univ. of Southern California, 3650 McClinton St., Los Angeles, CA 90089-1451, dibazar@usc.edu)

In this paper, we have proposed a new method of presenting a speech signal to the dynamic synapses neural network (DSNN) and optimal selection of input neurons for the speaker identification task. The DSNN developed by Liaw and Berger (1996) was employed to capture the speaker information from the action potentials, which was generated by wavelet filter banks. In order to sufficiently capture the speaker information, the time-delayed temporal patterns of action potentials are used as the inputs of the network. The optimal number of input neurons obtained by minimizing the total output error of the network while identifying a target speaker and rejecting the other unknown speakers. The genetic algorithm was employed to design a supervised learning rule for training of the network. The TI-46 dataset was used in both the training and testing phase. Comparative results with previous methods showed a better performance of the correct classification for the model-based dynamic synapses speaker identification system. [Work supported by DARPA and ONR.]

5aSC2. Discrete type dynamic synapses neural network: Speech recognition application. Hassan Heidari Namarvar (Dept. of Biomed. Eng., Univ. of Southern California, OHE-500, Los Angeles, CA 90089-1451, heidarin@usc.edu), Jim-Shih Liaw, and Theodore W. Berger (Univ. of Southern California, Los Angeles, CA 90089-1451)

A discrete type of the Dynamic Synapses Neural Network (DSNN) has been developed and applied to speech recognition. In order to speed up the training time of the network, a new discrete time implementation of the original DSNN [J.-S. Liaw and T. W. Berger, 1996] has been introduced based on the impulse invariant transformation. The new architecture of the network was trained with the Genetic Algorithms [H. H. Namarvar *et al.*, 2001] and tested against the continues-type DSNN. The overall speed of the new algorithm with discrete-time difference equation set is about 13 times faster than the same algorithm with the continuous differential equation set. This significant reduction of processing time not only decreases the training time but also makes the system better suited for real-time speech recognition tasks. [Work supported by DARPA.]

5aSC3. VUI aide to character recognition. Qiguang Lin, Mario Munich, and Kun Xia (VocalPoint Technologies, 847 Howard St., San Francisco, CA 94103)

Recognizing characters over the telephone is a challenging task, not only for computers but even for human beings. The reason is because there exist several highly confusable pairs/groups, e.g., the E-set, M/N, F/S, and A/J/K. Human beings resolve errors by saying phrases like: ''F as in Frank.'' In this report, a new character recognizer is first described, and then a novel voice-user-interface (VUI) is suggested to achieve better user experience with the task. The recognizer provides a hidden Markov model (HMM) for each character, augmented with two context-independent silence models. Furthermore, state-tying is made so that only discriminative states in individual HMMs are unique, while other states are shared. The next step is to design an effective VUI to facilitate character recognition. The VUI asks the user to say one character at a time. It then plays back the best result of recognition for confirmation. If the result is correct, the user proceeds; otherwise, the user says a keyword such as ''correction,'' and the recognizer retrieves an alternative candidate. Candidate selection can be made, for instance, according to an N-best list or a prior confusion matrix. It is found that the proposed VUI effectively disambiguates pairwise confusions such as F/S.

5aSC4. Tactical communications bring new challenges to automatic speech recognition. Thomas H. Crystal (Consultant, 2312 S. Joyce St., Arlington, VA 22202-2103, thcrystal@ieee.org), Astrid Schmidt-Nielsen, and Elaine Marsh (Naval Res. Lab., Washington, DC 20375-5337)

Automatic speaker-independent recognition of conversational speech has been making significant progress in recent years. Military communications challenge the robustness of current recognition systems. Compared to telephone conversations or news broadcasts, tactical communication has a reduced vocabulary. It consists of short utterances, limited to the task at hand, with occasional chat words. What makes recognition difficult is high levels of background noise in tactical environments (helicopters, tanks) and the degradation of the signal by military microphones, including noise canceling microphones. Of perhaps more importance, people alter their speech to overcome these degradations and the loss of intelligibility from communicating over vocoders. The DARPA-sponsored SPINE program is exploring these issues using speech from pairs of participants performing a collaborative task while communicating between separate sound booths. Each person sits in an accurately reproduced military background noise environment and uses the type of microphone and ear protection that would normally be found in that environment. The communication is conducted through vocoders. This presentation will review results from the SPINE evaluations of state-of-the-art speech recognition systems and from analyses of the transcripts in an attempt to understand how environmental factors and speaking changes affect recognizer performance. [Work supported by DARPA.]

5aSC5. A case study in optimizing TTS for Web content. Cynthia Hagstrom and Qiguang Lin (VocalPoint Technologies, 847 Howard St., San Franciso, CA 94103, hagstrom@vocalpoint.com)

This is a case study of linguistic optimization of the Speechify text-to-speech (TTS) engine for use in the VocalPoint voice-browser and applications. To browse a website by phone the TTS engine must be able to

handle two types of text, static system messages and dynamic content. The static messages include basic browser prompts such as the Default prompt: ''What would you like?'' or Help messages. In addition, the system must also handle dynamic content from the websites that are being browsed. The criteria of intelligibility, naturalness, and context-sensitivity were used to evaluate all the TTS output. Every static message was tested using a Speechify dialog tool. The TTS quality for dynamic content was evaluated through pronunciation bugs reported whilse browsing websites such as My Yahoo. Our study found that the text normalization module needs to be expanded to cover the vast variety of text on the Web. The study also suggested that three aditional functionalities would improve the caller experience: (1) make it possible to manipulate vowel length through punctuation or tagging; (2) make it possible to change intonation by manipulating pitch and phrase accents and tone breaks; and (3) add another phoneme to the sound inventory to allow flaps.

5aSC6. An introduction to the Buckeye corpus of conversational speech. William D. Raymond, Matthew J. Makashay, Robin Dautricourt, Keith Johnson (Dept. of Linguist., The Ohio State Univ., 222 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210, raymond@ling.ohio-state.edu), and Mark Pitt (Dept. of Psych., The Ohio State Univ., 222 Lazenby Hall, 1827 Neil Ave., Columbus, OH 43210)

Words in connected speech are subject to many types of phonological variation. Variation is both widespread and conditioned by an array of factors, including speech rate and context. However, little is known about the frequencies of variants or their acoustic realization. The variation-in-conversation (ViC) group at Ohio State has set about to quantify and measure the distributional properties of variants, with an eye toward their implications for spoken word perception. The current study reports on a corpus of approximately 40 h of conversation-style speech recorded during interviews of 43 midwestern speakers, stratified for age and sex. The speech was paired with text transcription, and is being phonetically transcribed using a combination of automatic and hand alignment. Details of the collection and transcription methodology will be presented, along with some parameters of variation that have emerged from the data. For example, in the corpus *t* is more likely to flap (16% of all underlying *t*'s) than *d* (8%). However, *d* is more likely to be deleted (36%) than *t* (18%), and almost all deletions of *d* are in the word final position (33%) unlike *t* (11%).

5aSC7. Concatenative text-to-speech synthesis based on waveform interpolation (a time frequency approach). Edmilson Morais and Grzegorz Dogil (IMS—Inst. of Natural Lang. Processing, Univ. of Stuttgart, AzenbergStrasse 12, D-70174, Stuttgart, Germany)

The time domain pitch synchronous overlap and add (TD-PSOLA) is the technique most used in comercial concatenative text-to-speech (TTS) synthesis systems. However, it is well known that TD-PSOLA presents several drawbacks. In order to overcome some drawbacks of the TD-PSOLA, this work presents a method based on time frequency interpolation (TFI) [Yair Shoham]. The method introduced here is a pitch-synchronous time-frequency approach of the waveform interpolation technique (WI) [Bastian Kleijn]. The goal of this work is to show that the TFI technique presents some important advantages to concatenative TTS synthesis. It allows pitch scale modification (PSM) independent of time scale modification (TSM) in a quite straightforward manner, and with high quality. TSM and PSM can be done in a continuous way, without any limitation of pitch period resolution. Moreover, the TFI technique allows simple, flexible, and efficient procedures to smooth diphone (or any other kind of unit) boundaries. The proposed system was evaluated using diphones and prosodies generated by the Festival system [Alan Black, Paul Taylor]. Subjective tests were performed, between the proposed TFI system and the standard TD-PSOLA system, highlighting the superior quality of the proposed system in comparison with TD-PSOLA.

5aSC8. Extraction and generation of aperiodic component in speech sounds. Hideki Kawahara (Dept. of Design Information Sci., Faculty of Systems Eng., Wakayama Univ., ATR/CREST, 930 Sakaedani, Wakayama 640-8510, Japan)

An effective method to extract and represent aperiodicity in speech sounds is proposed for making refinements on a high-quality speech analysis, modification and synthesis system STRAIGHT [Kawahara *et al.*, Speech Commun. **27**, 187–207 (1999)], based on time-frequency analyses by the use of an instantaneous frequency and group delay. The proposed signal representation consists of a frequency domain aperiodicity measure and a time domain energy concentration measure to represent source attributes. The frequency domain aperiodicity measure is defined as a ratio between the lower and upper smoothed spectral envelopes to represent the relative energy distribution of aperiodic components. The time domain measure is defined as an effective duration of the aperiodic component and event distribution pattern. These aperiodicity parameters and F_0 as time functions are used to generate the source signal for synthetic speech by controlling relative noise levels and the temporal envelope of the noise component of the mixed mode excitation signal, including fine timing and amplitude fluctuations. Relations between these parameters and temporal masking effects also will be discussed. [Work partly supported by MEXT Grant (C) 11650425 and CREST Japan.]

5aSC9. A novel method to estimate voice source and formant parameters based on an ARX speech production model. Takahiro Ohtsuka (Grad. School of Eng., Utsunomiya Univ.) and Hideki Kasuya (Faculty of Eng., Utsunomiya Univ., 7-1-2, Yoto, Utsunomiya 321-8585, Japan)

A novel method is presented to automatically estimate voice source and formant parameters from a speech utterance based on an autoregressive with exogenous input (ARX) speech production model. The method has been tested with both synthetic and natural speech materials and compared with the well-established linear prediction (LP) method in terms of the accuracy of estimated formant frequencies. The new method always revealed better performance than the LP method. In particular, the error of the proposed method was as small as 6% for the materials of an average pitch frequency of 447 Hz, while the LP method gave an error of 21%. The result clearly indicates superiority of the proposed method over the LP method for very high-pitched voices generated by females and children. Other key features of the proposed method include: (1) very low first formant frequencies of the high vowels /i/ and /u/ can be accurately estimated; (2) voicing source amplitude and open quotient of the glottal flow pulse are both reliably estimated; and (3) the implementation of an adaptive prefilter in the analysis remarkably improves the accuracy of estimated parameter values.

5aSC10. Complete optimization of excitation and model parameters in parametric speech coders. Khosrow Lashkari (DoCoMo USA Labs, San Jose, CA 95110)

In a speech coding system, synthesis error (the difference between the original speech at the encoder input and reproduced speech at the decoder output) is a more relevant measure of signal distortion than linear prediction (LP) error. By minimizing the synthesis error instead of the linear prediction error, the analysis and synthesis stages become more compatible. In a previous paper (J. Acoust. Soc. Am. **109**, 2491), an Analysis-by-Synthesis (AbS) technique for joint optimization of the excitation and filter parameters in parametric speech coders were presented. Using a gradient search in the root domain, synthesis error is minimized by reoptimizing the filter parameters for a given excitation. This paper provides a major improvement over the previous work, resulting in more than 3 dB of gain in the segmental signal-to-noise ratio. The improvement is due to a new algorithm for computing the gradient vector in the gradient search procedure. The paper describes the new algorithm and report on the im-

proved results. By adding an extra minimization step, this technique can be incorporated into the existing LPC, multipulse LPC, and CELP-type speech coding standards.

5aSC11. Estimation of vocal-tract area function from lip impulse response based on lossy vocal-tract digital filter model. Takemi Mochida and Masaaki Honda (NTT Commun. Sci. Labs., NTT Corp., 3-1 Morinosato-wakamiya, Atsugi-Shi, Kanagawa Pref. 243-0198, Japan, mochida@idea.brl.ntt.co.jp)

A method for estimating vocal-tract area function from lip impulse response based on lossy vocal-tract digital filter model is provided. Previous methods of lip impulse response [M. M. Sondhi and J. R. Resnick, **73**, 985–1002 (1983)] determines the area function from impulse response measured at lips by using the incident signal excited outside the oral cavity. This method is sensitive to the noise of measured impulse response. An alternative method [M. Honda and T. Mochida, Proc. 1st ETRW-SPM, 167–170 (1996)] is advantageous for accurate estimation under noisy acoustic condition, based on the vocal-tract digital filter model that optimizes the M sections of areas by using the impulse response longer than M samples. In this method, however, the vocal-tract model is based on the transmission line that treats the effect of loss in simplified form. The new method proposed in this study applies a mixture model of the lossless distributed circuit and the lumped circuit accounting for effects such as viscous loss, heat conduction loss, and loss due to wall vibration. The results on acoustic measurements have shown that the new method has better performance over the old method using optimization scheme based on simplified transmission line model. [Work supported by CREST, JST.]

5aSC12. The unwarranted effectiveness of smoothness as a criterion for estimating vocal-tract length from speech. Edward Neuburg (IDA-CCRP Thanet Rd., Princeton, NJ 08540)

Unambiguous quantitative characterization of speech sounds would be a useful thing. This talk will be about experiments done in an attempt to find such a characterization, by studying corpora of spoken vowel tokens all of which ''sound alike.'' Using Linear Predictive Coding (LPC) of order 10, 4 damped sinusoids (''formants'') were recovered from the tokens, and their bandwidths were algorithmically fixed as a function of frequency. (Other formant finders have also been tried.) Hypothesizing that a token is composed of 4 damped sinusoids is equivalent to hypothesizing that it was produced by a ''vocal tract'' with 8 sections; the shape (8 areas) of that tract is a function of its length, which is a free parameter of the formant-to-tract transformation. The experiments show that over a corpus of tokens of a given speech sound, simply choosing the vocal-tract length that produces the ''smoothest'' tract will make for shapes that are very nearly identical over all tokens, for all talkers (in particular for men and women), and are physiologically reasonable. Such 8-parameter shapes might be candidates for the desired characterization.

5aSC13. An audition-inspired method for producing high resolution spectrograms. Hilmi R. Dajani, Hans Kunov, and Willy Wong (Dept. of Elec. and Computer Eng., Univ. of Toronto, Toronto, ON M5S 1A1, Canada)

Inspired by the ability of the auditory system to dynamically track frequency and amplitude changes in modulated signals, an algorithm for high resolution time-varying spectral analysis is proposed. The main feature of the algorithm is an auditory-inspired filterbank followed by a peak detector. Mathematical analysis, with various types of synthetic modulated signals, demonstrates that the proposed method correctly demodulates these signals. The resulting time-frequency display, referred to as a fine structure spectrogram, shows the frequency and amplitude modulations in higher detail than is possible with conventional spectrograms. With recorded consonant–vowel syllables, the fine structure spectrogram detects small frequency and amplitude modulations in the initial glides and in the steady vowel formants. These modulations are not clearly seen in conventional time-frequency representations.

5aSC14. A nonstationary speech production model based on ARMAX analysis. Jun Huang and Stephen Levinson (Beckman Inst., Univ. of Illinois, Urbana, IL 61801)

Speech production is a very complicated process which includes both the linear propagative mode and the nonlinear flow mode. The plane wave propagation mode has been modeled by an autoregressive (AR) process and resulted in the widely used linear predictive (LP) model in speech production. In this work, we introduce an autoregressive moving average exogenous (ARMAX) model for speech production. We model the additional turbulence-induced acoustic sources and sinks by a nonstationary exogenenous noise process. First, we classify the English phonemes into five broad phonetic categories, namely, vowels, semi-vowels, stops, fricatives, and stops. Then we identify the model orders of each phonetic category from the acoustic observation. The model parameters are estimated by the Gauss–Newton algorithm. Finally, the prediction of the speech signal is performed through a recursive way after the model parameters of each phoneme are estimated. We tested this speech production model by synthesizing English phonemes and continuous sentences. Experimental results show that we can synthesize almost humanlike, natural quality speech signal based on this nonstationary speech production model.

Session 5aUW

Underwater Acoustics: Object Scattering and Imaging

Preston S. Wilson, Chair

Department of Aerospace and Mechanical Engineering, Boston University, Room 101, 110 Cummington Street, Boston, Massachusetts 02215

Chair's Introduction—8:25

Contributed Papers

8:30

5aUW1. Sub-critical insonification of buried spherical shells. Irena Veljkovic and Henrik Schmidt (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

The detection and classification of buried elastic targets presents a significant challenge to the shallow water mine countermeasures (MCM) community. Recent experimental and modeling studies have shown that in the case of subcritical isonification, significant evanescent field components can be converted into vertically propagating field components making the plane-wave, ray-tracing approach to the wave propagation inadequate. There has also been empirical evidence that the modal response of the target may be stronger than the direct backscattered return due to the evanescent nature of the sub-bottom field. The GOATS98 experiment had as one of its main objectives to explore the possibility of classifying features of the 3-D acoustics scattering by buried and proud objects using bistatic configurations such autonomous underwater vehicle (AUV) and horizontal line array (HLA). An unexpected physical phenomenon of a strong delayed flexural return of a flush buried spherical shell was observed in the AUV and the HLA data at frequencies much higher than predicted. A scenario similar to the GOATS98 experiment has also been modeled using OASES-3D target modeling framework to further investigate scattering mechanisms of the flush buried spherical shell under evanescent isonification, as well as to validate the modeling capability of the package. [Work supported by ONR.]

8:45

5aUW2. Bistatic synthetic aperture imaging of buried objects from an AUV. Kevin D. LePage (SACLANT Undersea Res. Ctr., Viale San Bartolomeo 400, 19138 La Spezia, Italy, lepage@saclantc.nato.int) and Henrik Schmidt (Dept. of Ocean Eng., MIT, 77 Massachusetts Ave., Cambridge, MA 01238, henrik@keel.mit.edu)

During SACLANTCEN's GOATS98 experiment an assortment of proud and buried objects were ensonified above and below critical angle by a parametric array and their scattered fields were observed bistatically on the nose array of a MIT Odyssey class AUV. Here the challenges of synthetic aperture sonar (SAS) processing for bistatic data sets are addressed and the detection of buried objects at subcritical ensonification angles is demonstrated. In addition, the differences between the measured resonant characteristics of spherical target scattering at below and above critical grazing angles is discussed.

9:00

5aUW3. Buried target classification with bistatic synthetic aperture sonar. Joseph R. Edwards and Henrik Schmidt (MIT, 77 Massachusetts Ave., 5-204, Cambridge, MA 02139, jre@mit.edu)

Buried target classification is of paramount importance in mine countermeasures (MCM) applications. Traditional methods use optimal detection techniques, e.g., monostatic synthetic aperture sonar (SAS) imaging, to classify objects by target shape. The shape of a target, however, does not necessarily indicate whether it is manmade (mine) or natural (rock), thus creating the possibility of an unacceptably high false alarm rate. Furthermore, the limited resolution available at the low frequencies required for seabed penetration may prevent the shape-based classification altogether, as the imaging wavelength is typically on the order of the target size. For these reasons, additional information about the character of the target is required. The current work investigates the use of frequency and aspect dependence of the target scattering for classification of canonical minelike targets, using a bistatic AUV-borne sonar receiver to interrogate spheres and cylinders in various burial conditions. Experimental evidence from GOATS98, an at-sea experiment performed off the coast of Elba Island, and supporting evidence from OASES-3D models are used to demonstrate clear distinctions between these canonical targets, and to illustrate the potential of a multi-static system for rapid and robust MCM. [Work supported by ONR and SACLANT.]

9:15

5aUW4. Impulse scattering processes of an elastic sphere. Christopher Feuillade (Naval Res. Lab., Stennis Space Center, MS 39529-5004)

The ''movable rigid sphere'' scattering model of Hickling and Wang [J. Acoust. Soc. Am. **39**, 276–279 (1966)] incorporates the translational ''rebound'' response of the scatterer to the incident field. However, as a physically unrealizable object, the rigid sphere has limited applicability. Recently, it was shown [J. Acoust. Soc. Am. **106**, 553–564 (1999)] that both the movable rigid sphere and Anderson's fluid sphere model [J. Acoust. Soc. Am. **22**, 426–431 (1950)] are limiting cases of Faran's elastic sphere scattering theory [J. Acoust. Soc. Am. **23**, 405–418 (1951)]. In this paper, the time domain impulse response is used to demonstrate the three dimensional scattering of an elastic sphere and its special cases. The critical factor is the shear rigidity modulus. For normal elastic values of the modulus, shear waves dominate the scattering response. When the rigidity is high, the shear waves travel more rapidly through the sphere, and its motion becomes increasingly characterized as a rebound translation without change of shape. Wave conversion at the boundary is also inhibited. For low or vanishing rigidity, only compressional waves are supported, and these determine the scattering response. The presentation will be illustrated by movie animations. [Work supported by ONR.]

9:30

5aUW5. Passive sonar calibration spheres. David M. Deveau (Naval Undersea Warfare Ctr. Det AUTEC, 801 Clematis St., West Palm Beach, FL 33401)

The need for calibrated sonar targets is addressed with the development and testing of a set of thin-walled spheres filled with a high-density fluid. Using historical research information as a guide, a set of thin-walled metal spheres was developed and filled with a high-density fluid. The combination of the spherical shape and the acoustic focusing effects of the fluid enhanced the acoustic scattering strength of the shape so that it was not only stable with temperature but also significantly greater in amplitude. The simple passive nature of the spheres makes them ideal acoustic

targets for long-term unattended deployments. Using a reference level measurement, each sphere was calibrated over a wide frequency range in order to provide the user with a curve of measured scattering strength vs frequency. The resulting curves showed a high degree of correlation between the individual spheres and the modeling that was used to extrapolate the theoretical values.

9:45

5aUW6. Target reconstruction from broadband acoustic backscatter. Quyen Q. Huynh, Jo E. Wilbur (Coastal Systems Station, Panama City, FL 32407), Nathan Intrator, and Nicola Neretti (Brown Univ., Providence, RI 02912)

Broadband acoustic backscatter data are examined for mine reconstruction using LFM and dolphin clicks. Broadband backscatter data from free-field and bottom mines were collected during tests conducted in ARL/UT's sand tank and sponsored by ONR. The data set includes multi-aspect broadband LFM (50–200 kHz) acoustic backscatter returns from a collection of targets. Both free-field and bottom target data were collected. Bottom target data sets include returns from two different levels of bottom roughness. Each data file is sampled over an interval that is large relative to the echo duration with sufficient sampling both prior to the echo onset and post-echo return to enable meaningful target characterization. Data collected over 360 deg in 2-deg increments are processed to generate a descriptive target pattern. The data are pulse compressed and system artifacts are removed using adjacent channels over the entire frequency band and within subbands. The target pattern is generated by applying the new adaptive wavelet filters to the scattered signals. In particular, the multi-aspect data that have several targets are tested to determine the ability of this new adaptive wavelet transform to reconstruct the external and internal structures of mines. Comparisons are made to previous results generated from broadband dolphin clicks.

10:00–10:15 Break

10:15

5aUW7. Sonar backscattering enhancements for small targets caused by caustics resulting from reflecting ocean surfaces. Ben Dzikowicz and Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

The reflection of sound by curved top or bottom ocean surfaces can produce caustics that enhance the sound level incident on a small target. Similarly, the scattering by the target reflected off a curved surface can be focused back onto the source transducer. These effects are simulated in a laboratory water tank experiment using a concave, cylindrical, Styrofoam surface and a small spherical target. The echo structure is simplest when the target is located at the cusp of the reflection caustic. A pulse can travel directly to the target and return or it can reflect off the surface on its way there, its way back, or both. For each reflection off the concave surface there is an observed enhancement so that the latest arriving pulse has a magnitude much larger than the earlier returns because it is doubly focused. The second return (though weaker than the doubly focused return) is also stronger than the direct echo. Application of a model for caustic wave fields [K. L. Williams *et al.*, J. Acoust. Soc. Am. **96**, 1687–1702 (1994)] enables simple predictions for the scaling of the relative amplitudes of the echoes with frequency. [Work supported by ONR.]

10:30

5aUW8. Broadband acoustic scattering from a bubbly liquid-filled compliant cylinder with known bubble population statistics. Preston S. Wilson, Ryan D. McCormick, Eun-Joo Park, Ronald A. Roy, and William M. Carey (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215)

In previous experiments, acoustic scattering from a bubbly liquid-filled latex tube was investigated in the 5–20 kHz frequency range. The mean bubble resonance frequency was 113 kHz, and a Wood limit effective medium theory was found to quantitatively describe the scattering throughout the measurement range, using a void fraction as the model's only fit parameter [Wilson *et al.*, Acoust. Res. Lett. Online (in press)]. New measurements will be presented that extend the frequency range to 150 kHz and include *in-situ* void fraction measurement. A preliminary analysis of the new data shows qualitative agreement with the effective medium theory up to about 35–40 kHz. Above this range, the qualitative trends predicted by the model are no longer apparent, and an increased ping-to-ping variability in the received echo is observed. As the frequency approaches individual bubble resonance frequency, one proceeds from a continuum response where the bubble distribution is not important to a region where individual bubbles and groups of bubbles play a role. Moreover, as the frequency goes up, losses in the latex rubber tube wall should become increasingly important. An analysis of the data, which includes bubble resonance effects and tube wall losses, will be presented. [Work supported by ONR.]

10:45

5aUW9. The three-dimensional Fredholm equation of the first kind and its application to the inverse problem of submerged objects in a waveguide. M. F. Werby (NRL Code 7180, Stennis Space Center, MS 39529 and Dept. of Phys., Catholic Univ. of America, Washington, DC 20064)

Assume one has knowledge of the Green's function in some ocean waveguide and a measurable signal scattered from a source. This may be treated by solving the Fredholm equation of the first kind (FEFK). The task is to determine any features from the measured field by solving the FEFK. We use synthetic data and produce measurents and determine what information may be recovered by solving this equation.

11:00

5aUW10. Do submerged objects twinkle from gated pulse signals in turbulent wave guides? Michael Werby (NRL Code 7180, Stennis Space Center, MS 39529 and Dept. of Phys., Catholic Univ. of America, Washington, DC 20064)

If one illuminates the ocean with a time-gated acoustical signal will an inclusion such as an elastic shell insonfied by the pulsed twinkle or, in the jargon of some researchers, will it scintillate? Twinkling from stars owes its origin to a mildly turbulent atmosphere. The ocean is often much more turbulent and we examine the distortions that one gets from scattered pulse-gated signals.

11:15

5aUW11. More on the nonhomogeneous Sturm–Louiville problem and its application to inclusions in a waveguide and the extraction of target properties. M. F. Werby (NRL Code 7180, Stennis Space Center, MS 39529 and Dept. of Phys., Catholic Univ. of America, Washington, DC 20064)

The second order solution of the wave equation is useful in adopting a method for solving the related inhomogeneous version in which the inhomogeneous term comes from a known target or an unknown object in the waveguide. Formulations are presented for these cases with examples.